STRUCTURAL STUDY OF LIPID-BINDING PROTEINS

A Dissertation

by

HAN-CHUN TSAI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	James C. Sacchettini
Committee Members,	Karen L. Wooley
	Thomas R. Ioerger
	David H. Russell
Head of Department,	David H. Russell

August 2013

Major Subject: Chemistry

ABSTRACT

Tuberculosis and malaria are among the most deadly infectious diseases in the world. The prevalence in regions without well-established public health causes economical and financial burdens for both society and patients. There is an urgent need to find effective treatments due to the emergence of drug-resistant strains. The aim of the studies reported here was to gain knowledge from the protein structures that can lead to the elimination of these pathogens. In these studies, protein crystallography is the main method used to solve protein structure. Based on the protein structure, we used different methods to characterize the protein function of three lipid-binding proteins (LprG, LprA, and gp232), and to identify potent inhibitors against *Plasmodium falciparum* enoyl-ACP reductase (PfENR), a drug target protein involved in central lipid metabolism. To characterize the function of two lipid-binding proteins (LprG and LprA), liquid chromatography-mass spectrometry (LC-MS) was used to analyze the ligand extract. In the study of tail fiber protein from mycobacteriophage, we used protein sequence alignment to identify gp232 as a major tail fiber protein, which potentially binds to lipids on the cellular surface of mycobacteria. A pull-down assay and imaging methods (fluorescence microscopy and electron microscopy) were conducted to confirm the function of gp232. In the structural study of PfENR, the structure-activity relationships method was used to find potent inhibitors against PfENR, which would show stronger

inhibition than the known inhibitor triclosan. The triclosan-like analogs with modification at the 5-position revealed a new binding site in PfENR that has great potential for improving the potency of inhibition. We found that two inhibitors containing the core structure of piperidine and tetrahydroquinoline reached this new binding site and were 10-fold more potent than triclosan. The structural study of PfENR provides structural insights into the inhibitor-binding site that can lead to the discovery of new drugs. The comprehensive knowledge that we gained from the structural studies of these lipid-binding proteins provide new information that could lead to a greater understanding of pathogen physiology or guide the discovery of effective treatments to eliminate the pathogens.

TABLE OF CONTENTS

Page

ABSTRACT ii

TABLE OF CONTENTS iv

LIST OF FIGURES vi

LIST OF TABLES x

1. INTRODUCTION 1

    1.1 The functions of lipids in living organisms 2
    1.2 The important functions of lipids in mycobacteria 3
    1.3 The current epidemic and background of tuberculosis 6
    1.4 Three major types of lipids in mycobacteria 12
    1.5 Mycobacterial lipid-associated proteins 19
    1.6 Application of structural biology 28

2. STRUCTURAL STUDY OF LPRA AND LPRG 31

    2.1 Introduction 31
    2.2 Results and discussion 36
    2.3 Methods 72

3. STRUCTURAL ANALYSIS OF PFENR FOR INHIBITOR OPTIMIZATION 78

    3.1 Introduction 78
    3.2 Results and discussion 83
    3.3 Methods 122

4. FUNCTIONAL IDENTIFICATION AND STRUCTURAL STUDY OF TAIL FIBER PROTEIN FROM MYCOBACTERIOPHAGE BXZ1 125

    4.1 Introduction 125
    4.2 Results and discussion 129
    4.3 Methods 163

Page

5. CONCLUSION 170

REFERENCES 174

APPENDIX A THE CRYSTAL STRUCTURE OF LPRG 187

APPENDIX B THE STRUCTURAL STUDY OF TRICLOSAN DERIVATIVES WITHMODIFICATION AT THE 2'-POSITION 197

APPENDIX C THE STRUCTURAL STUDY OF TRICLOSAN DERIVATIVES WITH MODIFICATION AT THE 5-POSITION 205

LIST OF FIGURES

Page

Figure 1.1 Chemical structure of triacylglycerol (TAG)..... 4

Figure 1.2 a Electron micrograph of *Mycobacterium tuberculosis* 11

Figure 1.2 b Model of mycobacterial membrane and cell wall..... 11

Figure 1.3 Chemical structures of phospholipids 16

Figure 1.4 Major components in the cell wall 18

Figure 1.5 Diagram of fatty acid biosynthesis pathway 23

Figure 1.6 Lipoarabinomannans (LAM) biosynthesis pathway 26

Figure 2.1 a Superimposition of two crystal structures of LprA (molecules A and B) from an asymmetrical unit cell..... 38

Figure 2.1 b Overview of the LprA surface 39

Figure 2.2 a Mass spectral analysis of the ligand extraction from *E. coli* purified LprA..... 43

Figure 2.2 b LC-MS profile of the *E. coli* purified LprA 46

Figure 2.2 c MS-MS of the bound ligands from the extraction 48

Figure 2.3 a Crystal structure of LprA with PG (16:0, 17:1) in the cavity 49

Figure 2.3 b Crystal structure of LprA with PG (16:0, 16:1) in the cavity..... 50

Figure 2.4 a Identification of potential ligands from the Mtb lysate by LC-MS 54

Figure 2.4 b Differential LC-MS profiles of *E. coli* purified LprA and the Mtb lysate incubated LprA from methanol extraction 56

Figure 2.4 c Differential LC-MS profile of the *E. coli* purified LprA and the Mtb lysate incubated LprA from ethyl acetate extraction 58

Figure 2.5 LC-MS profiles for analyzing the extraction of the full-length LprA purified from *M. smeg* 61

Figure 2.6 Structural comparison of the cavity in LprA and LprG shows a major difference at helix4 68

Figure 3.1 Overall PfENR structure 85

Figure 3.2 Active site of PfENR including the triclosan- and NAD^+-binding site ... 87

Figure 3.3 a Overall view of crotonoyl-CoA binding in the active site (upper) and the structure of crotonoyl-CoA (lower) 89

Figure 3.3 b Surface view of PfENR with crotonoyl-CoA 91

Figure 3.3 c Model of ACP (PDB:4ETW) docking on PfENR 91

Figure 3.3 d Possible residues making contact in PfENR and ACP 92

Figure 3.4 a Crystal structure of PfENR with triclosan analogs of the 6-position substitution: compound **12** 96

Figure 3.4 b Compound **14** 97

Figure 3.4 c Compound **23** 97

Figure 3.4 d Compound **43** 98

Figure 3.4 e Overlap of 6-subsitution analogs binding in the flipped orientation 98

Figure 3.4 f Compound **32** 99

Figure 3.4 g Ring A of the compound **32** blocks NAD^+ binding 99

Figure 3.5 Binding of the piperidine compound **325** in the active site 107

Figure 3.6 a Crystal structure of indole compound binding in the active site: indole compound **313** 110

Figure 3.6 b Indole compound **059** 110

Figure 3.6 c Indole compound **160** 111

Figure 3.6 d Indole compound **813** 111

Figure 3.6 e Indole compound **023** 112

Figure 3.6 f Indole compound **363** 112

Figure 3.6 g Indole compound **865** 113

Figure 3.6 h Indole compound **914** 113

Figure 3.6 i Indole compounds overlap 114

Figure 3.7 Crystal structures of the THQ compounds binding in the active site 117

Figure 3.7 a THQ compound **698** 117

Figure 3.7 b THQ compound **853** 118

Figure 3.7 c THQ compound **855** 118

Figure 3.7 d THQ compound **987** 119

Figure 3.7 e THQ compound **696** 119

Figure 3.7 f THQ compound **745** 120

Figure 4.1 Gene organization of tail assembly in (a) TM4 and (b) bxz1 131

Figure 4.2 Sequence alignment between the major TFP and the minor TFP from two strains of *Myoviridae* 133

Figure 4.3 Truncated gp232 with N-terminal GFP tag binds to *M. smegmatis* as observed by fluorescence microscopy 136

Figure 4.4 Truncated gp232 labeled with Ni-NTA-nanogold binds to the bacterial surface of *M. Smeg* as viewed by electron microscopy (JEOL1200) 138

Figure 4.5 Truncated gp232 binds to the cell debris of *M. smegmatis* 140

Figure 4.6 Overall structure of the truncated gp232 143

Figure 4.7 Overall outlook of the connection head domain in the truncated gp232 . 145

Figure 4.8 Shaft domain of the truncated gp232. 147

Figure 4.9 Foot domain of the truncated gp232 149

Figure 4.10 Structure of tail fiber proteins from bacteriophages and viruses 152

Figure 4.11 Potential receptor-recognition surface of the bacteriophage TFPs and the viral fiber proteins. 156

Figure 4.12 A mushroom-like cavity at the center of potential receptor-recognition surface 158

LIST OF TABLES

Page

Table 2.1 Measurement of the dissociation constant 64

Table 3.1 IC_{50} of the 6-substitution triclosan analogs 94

Table 3.2 Chemical structure and measured IC_{50} of the compounds with the new backbone 104

1. INTRODUCTION

Lipids are essential components for living organisms from bacteria to mammals. This essentiality is supported by the functions of lipids in cells that intrigue us to understand lipid metabolism that is related to drug target and pathogen physiology. Because they are associated with lipids, lipid-binding proteins play an important role in carrying out the function of these lipids. In addition, some lipids are signals used for recognition by the host immune system to identify the intrusion of pathogens. This recognition of lipids on the cell wall of pathogen also is used by phages and is performed by a type of proteins from the phages. The proteins that are associated with bacterial recognition are involved in the phage infection specificity, and they show great potential in therapeutic applications such as targeted drug delivery systems. To find effective treatments for deadly infectious diseases such as tuberculosis (TB) and malaria, structural studies of lipid-binding proteins can guide us from a structural perspective to design potent drugs, understand the pathogen physiology, and improve the efficacy of known treatments. Therefore, in this study, I am interested in the protein structures of four potential lipid-binding proteins that conform to the requirements and functions mentioned above.

1.1 The functions of lipids in living organisms

Lipids comprise a large number of naturally occurring cellular components and are important for the survival of all living organisms [1, 2]. They are defined by their hydrophobicity, which distinguishes them into a group of biomolecules that is different from other biomolecules. The atomic composition of lipids is carbon, hydrogen, and oxygen, and some forms of lipids contain phosphorus, nitrogen, and other elements. They are essential components in all organisms and constitute a broad group of naturally present molecules with chemical modifications that assist in the survival of organisms. For example, fats and oils, including diverse fatty acids and cholesterol, are lipid forms that have metabolic and nutritional functions that support life [3]. The structural functions of lipids are essential for all organisms, from single-cell to higher organisms, because the formation of cell membranes requires phospholipids [4]. Some lipids such as eicosanoid serve a role in the intercellular signaling for inflammation in humans [5]. Therefore, lipids are instrumental in the evolution of species, playing important structural roles in building cellular membranes, acting as primary source of energy (9 kcal/g), and functioning as signaling elements or the precursors of hormones.

Lipid functions are indispensable for pathogens to survive inside the host. Lipids provide protection or a means that modulates the host immune response [6]. To survive and cause infection, pathogens have to propagate through replication, which requires the production of large amounts of lipids to make the cellular membrane. Due to the lipid dependency of the pathogens residing in the host, studies related to lipid metabolic, structural, or signaling functions would help to elucidate pathogen survival mechanisms.

Then, therapeutic treatments could be designed based on the survival mechanisms. For example, the essentiality of lipid production in building cell membrane has led to the design of inhibitors that could block lipid biosynthesis. A front line anti-tuberculosis drug, isoniazid, inhibits the enzyme, InhA, which is involved in fatty acid biosynthesis [7, 8]. Moreover, the knowledge of structural and signaling functions could guide us to develop applications in diagnosis and disease prevention.

1.2 The important functions of lipids to mycobacteria

Energy storage

Lipids such as fatty acids and triacylglycerol (TAG) (Fig 1.1), which are more accessible for catabolic reactions compared to smaller units of carbon building blocks (acetyl CoA), makes them the ideal storage of energy[9]. These lipids supply energy through respiration and provide a carbon source[10]. When a harsh environment lacks sufficient nutrients, such as inside the macrophage, mycobacteria will enter a dormant (non-replicative) state that causes a re-routing of carbon flow in the tricarboxylic acid (TCA) cycle [8]. For aerobic organisms, the TCA cycle is the key component in the metabolic pathway that generates energy from carbohydrates or sugars. The metabolic flexibility in mycobacteria allows it to be persistent in the macrophage by switching its carbon source from sugars to fatty acids. Therefore, fatty acids, which are broken down by beta-oxidation and used to generate acetyl CoA for the TCA cycle, become the main energy source for mycobacteria to survive in energy-limited conditions[11]. Due to this capacity of the fatty acids stored in TAGs, mycobacteria accumulate TAG-based lipids

from the macrophage in which mycobacteria are engulfed by the first-line of host defense [12, 13]. This is also supported by the fact that large amounts of TAGs are observed in the sputum of tuberculosis patients [13].

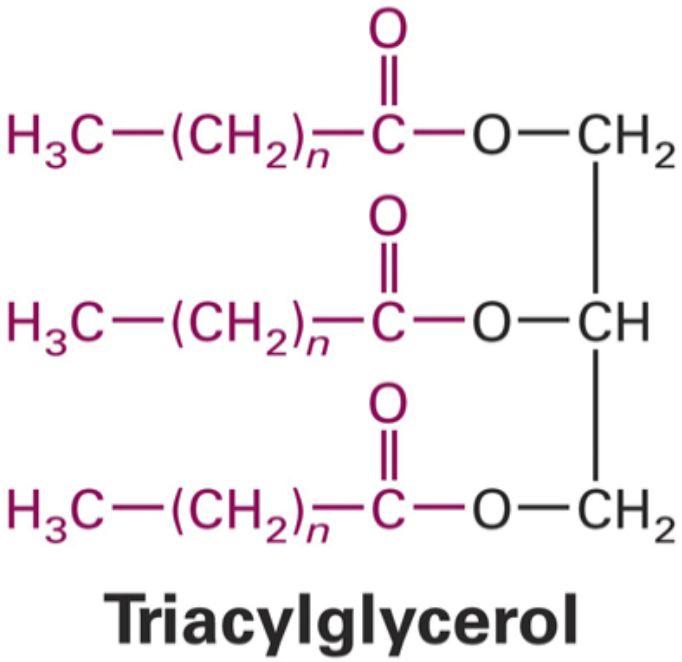


Figure 1.1 Chemical structure of triacylglycerol (TAG) [14]

Structural function and protection

A few classes of lipids such as phospholipids have a unique physical property in that they possess amphiphilic characteristics. These lipids contain a polar head group and a non-polar tail group, and they are the main components that make the double layers of cellular membranes. Not all the lipids used for building structural compartments are amphipathic, but they all act as the structural components for maintaining the integrity of the cell membrane[15]. This membrane builds a compartment to isolate its own cellular contents from the surrounding environment, and provides a selectively permeable barrier for protection[16]. Therefore, making the structural compartment is essential for mycobacteria to multiply in the host and lead to further infection. The main components, phospholipids, are required for pathogen replication once they emerge from the dormant state[16, 17]. To confer protection, cell membranes must be strong and stable enough to maintain cellular integrity, and flexible enough to bear conformational or environmental changes. The composition of lipids in the membrane is dynamic in response to a wide range of environment conditions (e.g., temperature, pH, and the hostile environment inside the host phagosomes)[16]. Changes in the membrane lipid composition of single-celled organisms have a more direct impact than those in multicellular organisms. For example, the acyl chains of phospholipids determine the viscosity of the membrane and influence cell permeability. Disruption of the acyl chain modification on phospholipids leads to a loss of stability of the membrane, which in turn causes single-celled organisms such as bacteria to become weak and susceptible to stress such as antibiotics[16]. However, mycobacteria have unique cell wall structures that make them waxy and hard

to penetrate. The constituents of the cell wall are mainly mycolic acids that have long carbon chains with hydroxyl or branched modifications, which are normally metabolic burdens in nutrient-restricted environments. By contrast, the same mycolic acids in the outer shell of the cell wall are the reason that several strains of mycobacteria are successful pathogens that co-evolve with humans and develop resistance to treatments[18, 19].

Signaling function

Some pathogens such as *E. coli*, *Mycobacterium tuberculosis* (Mtb), and *Plasmodium* take advantage of certain lipids that have an ability to modulate the host immune system for survival in the host[19, 20]. For mycobacteria to cause infection and proceed into a disease state after invasion, modulation of immune signaling in the host is one of the key factors to survive and propagate under the harsh environment that is created by the host immune system[21].

1.3 The current epidemic and background of tuberculosis

Tuberculosis (TB) is an infectious disease among higher animals including humans and is caused by the pathogen, *Mycobacterium tuberculosis*. TB typically affects the lung (pulmonary TB), but can also infect other organs in the human body such as skeletal TB and central nervous system TB. Pulmonary TB has a long history since the seventeenth century, and the infectious nature of the disease was reported at that time. Later, Benjamin Marten, an English physician, published that the disease is caused by

microorganisms and is transmitted by close contact[22]. In 1882, Robert Koch discovered that the symptoms and the disease are caused by a single pathogen, which was later named *Mycobacterium tuberculosis*.[22] The effort to find treatments for TB has been explored with three approaches: strengthening immunity, the BCG vaccination, and the use of chemotherapy. In the early 1990s, the incidence of TB cases increased rapidly due to co-infection in HIV-positive patients[22]. In 1997, the World Health Organization (WHO) started to monitor TB epidemiology globally, and kept optimizing the strategy for the administration of treatment. Despite all the efforts in combating TB, there were 8.7 million new cases in 2011 and 1.4 million TB deaths (990,000 deaths of patients without HIV-infection and 430,000 deaths of HIV-positive patients)[23]. The global burden of TB is enormous; geographically, the biggest burden is in Asia and Africa, which had the largest number of incident cases in 2011. Among The rate of the overall global incident cases is 59% in Asia and 26% in Africa[23]. The estimated TB incidence rate of all six regions, including Africa, Europe, South-East Asia, Western Pacific, The Americas, and Eastern Mediterranean, are declining. In addition to monitoring the incidence numbers, the WHO surveillance of TB care and control evaluates the global prevalence and the mortality rate. The prevalence was estimated as 12 million people in 2011, which translates to 170 cases per 100,000 people, and this is 36% lower than the prevalence rate in 1990. The global TB mortality rate shows a decline of 41% since 1990[23].

Without chemotherapy, TB mortality rates are high. When the TB infection progresses, and patients without co-infection of HIV are positive in the sputum-smear

test, 70% of all patients die within 10 years. However, 90% of TB cases can be cured using the short-course regimen consisting of front-line drugs that have been available since the 1980s[23]. After patients are diagnosed by the sputum-smear microscopy test or the culture-method test, applying a proper treatment regimen is crucial to fight off the infection. Therefore, DOTS (directly observed treatment, short-course) is implemented to reinforce the patient's compliance. Due to the unique nature of Mtb, such as the thick layer of the cell wall and the complex metabolism (with the ability to switch to the non-replication state), the effective treatment regimen requires patients with newly diagnosed drug-susceptible TB to take a 6-month course of anti-tuberculosis drugs. This treatment regimen includes four front-line drugs including isoniazid, rifampicin, ethambutol, and pyrazinamide. The implementation of DOTS requires physical monitoring of patients taking isoniazid-rifampicin-ethambutol-pyrazinamide for 2 months, and subsequently isoniazid-rifampicin for 4 months. This program shows a success rate higher than 95%[23]. The administered DOTS strategy is aimed to prevent the emergence of drug-resistant strains of Mtb.

Two variables that influence disease control

The co-infection of TB with HIV and the occurrence of drug-resistant TB are two variables that exacerbate the TB epidemic. The emergence of the AIDS pandemic was widely recognized in the 1980s[24]. The major increase in the number of cases and the mortality rate of TB persisted in southern and east Africa from 1990 to 2004. HIV-positive patients with an impaired immune system who are also infected with TB have a

higher risk of developing TB. The WHO report in 2011 states that 13% of people who developed TB are HIV-positive[23]. Due to the compromised immunity, HIV infection also affects the TB cure rate and mortality rate. The other factor that influences disease control is the identification of drug-resistant strains of Mtb. Drug-resistant TB refers to an infection that cannot be controlled by the first-line drugs. For example, the multidrug-resistant (MDR) Mtb is resistant to at least two front-line drugs such as isoniazid and rifampicin. Among all the new cases of TB, 3.7% are estimated to be MDR-TB cases[23]. The incidence rate of MDR-TB is high in Eastern Europe and Central Asia. The standard treatment for patients infected with MDR-TB lasts 20 months or longer, and the chemotherapy drugs are more toxic to humans. Due to the lack of potent drugs for MDR-TB, the WHO launched drug-resistance surveillance in 1994 to monitor the epidemic of MDR-TB. This also revealed the need for new treatments for TB to prevent an upsurge in the incidence number.

Characteristics and cell structure of Mtb

Mycobacterium tuberculosis is a pathogenic bacterial species. The unique cell structure has characteristics of both Gram-positive and Gram-negative bacteria. The shape of mycobacteria is thin-rod-like, with a length of 2–5 μm and width of 0.2–0.3 μm (Fig 1.2a)[25, 26]. Mtb has a slow growth rate and divides every 15–20 hours. As an aerobic bacterium, the growth rate strictly depends on the available oxygen[22, 27]. When Mtb is engulfed by macrophages, it stops replicating and stays in a dormant state due to the limited nutrient conditions. Under the protection of the thick cell wall, Mtb is

able to avoid eradication by the host immune system.[28, 29] The mycobacterial cell wall reveals a unique structure of Gram-positive bacteria, with an extra outer membrane consisting of high lipid content; with a thickness of 7−8 nm, the cell wall is similar to the outer membrane of Gram-negative bacteria (Fig 1.2b)[25, 30]. The inner lipid bilayer mainly contains phospholipids, which separate the cytoplasmic region from the periplasm. The periplasmic region is the compartment formed between the outer cell wall and the inner membrane, which is a unique feature of the mycobacteria. This compartment is built by polysaccharides, such as peptidoglycan (PG) and arabinogalactan (AG)[25]. The waxy outer cell wall is composed of mycolic acids attached to the AG moieties, and provides a barrier and a habitat favorable for all the outer glycolipids to form a cell envelope. These glycolipids are shown to act as a signal that modulates the host immune response[28].

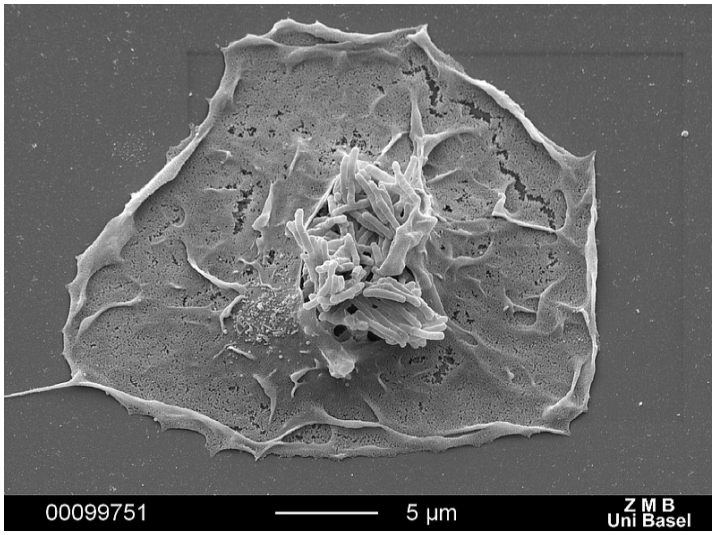


Figure 1.2a Electron micrograph of *Mycobacterium tuberculosis*[31]

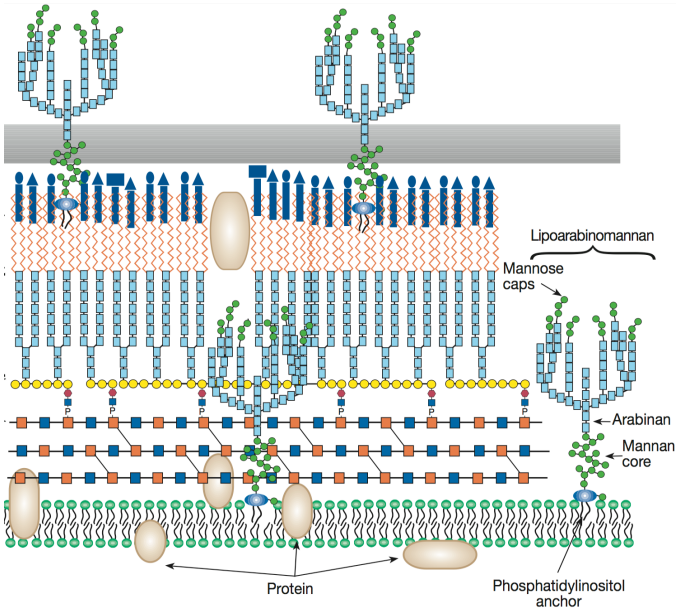


Figure 1.2b Model of mycobacterial membrane and cell wall [29]

Defense mechanism

Once mycobacteria invade a host, the host front-line protection is the innate immune response, which involves antigen-presenting cells (APC) like macrophages and dendritic cells. Macrophages engulf the mycobacteria and become phagosomes, where the cellular environment becomes acidic and nutrient-limited. While mycobacteria are lysing in the phagosome, the antigen-presenting cells present the pathogen receptor to T-cells. This is the adaptive immune response that recruits CD4 and CD8 T-cells to eliminate mycobacteria[32].

1.4 Three major types of lipids in mycobacteria

Lipids for the storage of energy and carbon

Fatty acids are the reserves of energy and carbon storage and serve as metabolic precursors for other lipid biosynthesis pathways such as mycolic acid and phospholipids[25] . The chemical structure is formed by more than four carbons in a long chain, which is attached to a carboxylic acid group, and the acyl chain length can be in the range of 4–20 carbons (C4–C20)[11]. The covalent bonds between carbons in fatty acids are either a single (saturated) or double (unsaturated) bond. There are other modifications on the acyl chain of fatty acids. The common modifications are methylation and cyclopropantion, which are used to alter the cellular membrane viscosity[16, 33]. For example, palmitic acid (C16:0) is linear and can pack very tightly, which leads to poor fluidity and lower permeability into the membrane. The significance of these components with long or short carbon-chain lengths to the Mtb pathogenicity or

adaptation to the environment has not been fully established. However, from monitoring the changing lipid composition in the bacterial membrane, the composition of the fatty acid carbon chain length seems to be optimized by evolution of bacterial strains that change to adapt to the temperature, culture age, and nutrients[16, 34]. In general, the most abundant fatty acids in Mtb are palmitic acid (C16:0) and tuberculostearic acid (C19:0)[35]. During culture aging, the tuberculostearic acid (C19:0) composition of the lipids increases in all species, whereas the composition of oleic acid (C18:0) and palmitoleic acid (C16:0) decreases[36, 37]. Mycobacterial species with different growth rates show different chain lengths in the fatty acid composition. The normal population of fatty acids is C16–C19[38]. Despite a lack of understanding of the correlation between fatty acid composition and bacterial viability, these fatty acids are the essential building blocks of phospholipids, glycolipids, triacylglycerides, and other cell wall components (such as mycolic acids) through further metabolic and modification reactions.

Lipids and the structure of the plasma membrane

Phospholipids are the major elements of cell membranes[11]. A phospholipid molecule is constructed with four components (Fig 1.3), including fatty acid, a platform for fatty acid attachment, a phosphate, and a head group attached to the phosphate. All phospholipids contain a diglyceride that has two fatty acyl chains connected to the glycerol moiety at the *sn*-1 and the *sn*-2 positions. A phosphate moiety is esterified to the glycerol at the *sn*-3 position. This is the basic unit of phospholipid, also known as

phosphatidic acid. Only small amounts of phosphatidic acid are present in membranes[39]. However, the molecule is a key intermediate in the biosynthesis of other phospholipids. The forth component, the head group, is hydrophilic, and differences in the head groups are important criteria that modify the property of the phospholipids. The hydrophobic tail and the hydrophilic head group result in phospholipids having an amphipathic character. The fatty acyl chains in the tail are repelled by water and are forced to aggregate, and this leads to the formation of the cell membrane lipid bilayers. The fatty acid components provide a hydrophobic barrier, whereas the remainder of the molecule has a hydrophilic character that enables interaction with the environment.

Based on the head group, phospholipids are classified into eight types: phosphatidic acid (PA), phosphatidylglycerol (PG), phosphatidylethanolamine (PE), phosphatidylserine (PS), phosphatidylcholine (PC), phosphatidylinositol (PI), cardiolipin (CL), and sphingomyelin (Fig 1.3). Bacterial cell membranes, in general, have five phospholipids, which are PA, PG, PE, CL and PS. However, mycobacterial cell membrane contains one extra type of phospholipid, phosphatidylinositol (PI), which is not observed in most bacteria, except for the species in the suborder of *Corynebacterineae[39]*. Among all the phospholipids in the mycobacterial cell membrane, PG, PE, and CL are the most abundant components, followed by PI[11]. However, the percentage composition of these phospholipids changes in response to environmental or cellular changes[39]. The distribution of these phospholipids is heterogeneous. The hypothesis proposed for lipid polymorphism was that the membrane

formed by phospholipid was required to be dynamic to produce a specific environment for recruiting membrane proteins to function. The environment is defined by the potential, which is determined by the polarity of the modification on phosphatidic acid[16]. For example, the ethanolamine moiety of PE confers a positive charge on the cell membrane, whereas the phosphate moiety of phospholipids produces a negative charge on the surface of the membrane. The other head groups likely lead to different polarity. In addition, the fatty acyl chains of the phospholipids can be modified to adjust the fluidity and the rigidity of the membrane by adding cyclopropanyl groups or methyl groups and unsaturations in the acyl chains[16, 33].

The function of phospholipids is more than building barriers that protect cells from the environment. Some phospholipids in pathogenic mycobacteria serve as a signal to mediate bacterial stress responses or host immune responses. Phosphatidylinositol 3-phosphate (PI3P) is identified in both *Mycobacterium smegmatis* and *Corynebacterium glutamicum* and only shows a transient accumulation when the salt concentration increases rapidly[40]. The PI3P could serve in the signaling cascade to react to stress from the environment. Another example is observed in *Plasmodium*, the cause of malaria. This pathogen can survive under constant monitoring from the host immune system. Humans get *Plasmodium* infection through a mosquito bite. At one stage of its life cycle, the pathogen dwells dormant in hepatocytes (liver cells) after drifting in the blood stream. Before entering the next infection stage by invading erythrocytes (blood cells), the parasites propagate in hepatocytes. The dormant *Plasmodium* in liver survives under the surveillance of the host immune system by inhibiting the exposure of

phosphatidylserine on the outer leaflet of host plasma membrane[41]. *Plasmodium* utilizes this change of phospholipid composition in the membrane to suppress the signal to the immune response and thereby prevent elimination.

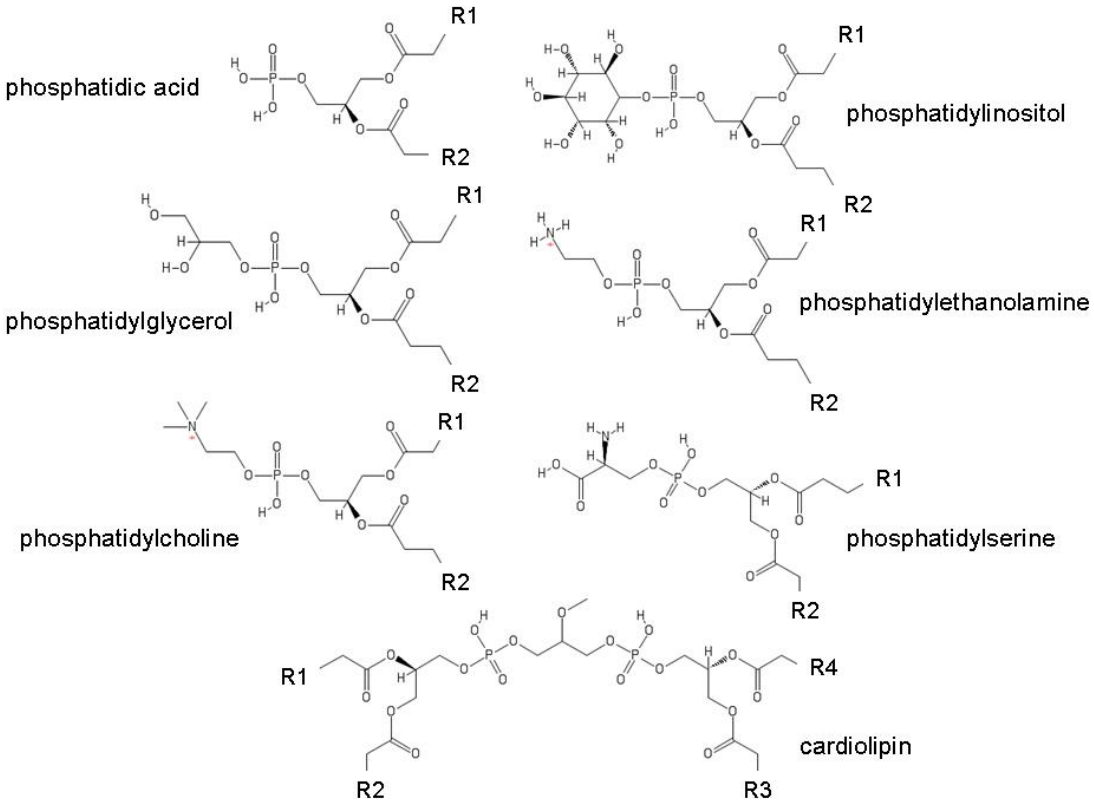


Figure 1.3 Chemical structures of phospholipids

Lipids and the structure of the cell wall

The mycobacterial cell wall is waxy and provides a second layer of protection built in a layer outside the plasma membrane. It is essential for the growth and survival of the pathogen inside the infected host because of its low permeability. The cell wall consists of two segments, the core and outer segments (Fig 1.4). The core layer of the

cell wall is comprised of three covalently linked macromolecules beyond the membrane. These three major chemical constituents are PG, AG, and mycolic acids[18, 42]. The PG is covalently attached to the AG component, which is linked sequentially to the mycolic acid with their long meromycolate chains and short alpha chains. This layer is a covalently fused structure of the three macromolecules, also known as the mycolyl-arabinogalactan-peptidoglycan (mAGP) complex, and is the core of the cell wall. The outer segment consists of free lipids, including some fatty acids with modifications (alpha chains), and the cell-wall associated proteins[18]. The free lipids, which are non-covalently associated with the mAGP complex, are mainly fatty acyl chains with modifications in the sugar moiety. These glycolipids are commonly found in mycobacteria and include phosphatidylinositol mannosides (PIMs), the phthiocerol-containing lipids (PDIM, DIM), lipomannan (LM), sulfolipids, cord factor, and lipoarabinomannan (LAM)[43]. The sugar modification makes these glycolipids become soluble in various solvents, but the mAGP complex remains insoluble. This unique cell wall structure of mycobacteria with various lipid contents is crucial for pathogenicity when mycobacteria proceed in the infection process. The proteins on the surface of the cell wall and the polar molecules of free lipids have the capability of cellular signaling, and therefore become effecter molecules in the pathogenicity and survival in the defense system of the host[6]. The waxy mAGP complex provides low permeability characteristics that contribute to the resistance of the pathogen to any common chemotherapeutic agents[18].

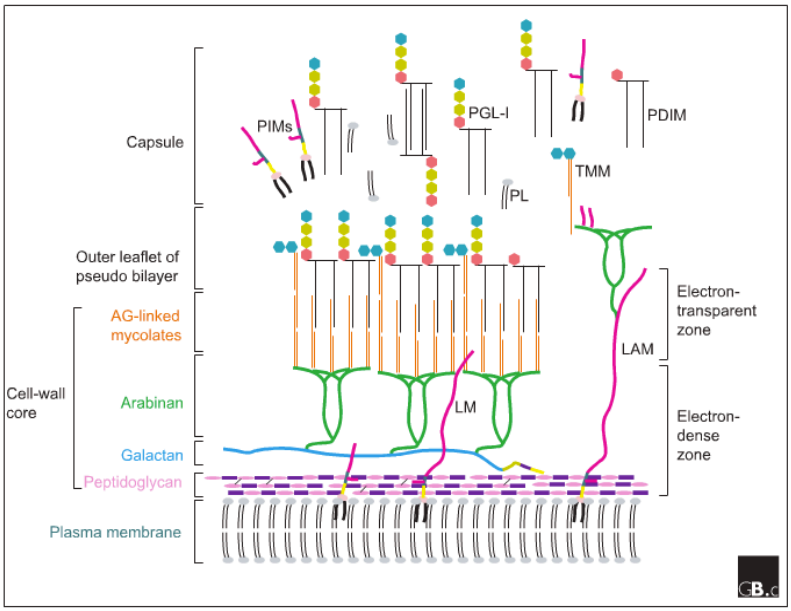


Figure 1.4 Major components in the cell wall[42]

One of the major lipids in the mycobacterial cell envelope is the phosphatidylinositol-derived lipoglycan family of LAM/LM/PIM (Fig 1.4)[39]. These glycolipids share a common core structure of mannoyl-phosphatidyl-*myo*-inositol (MPI), with phosphatidylinositol as the base. The PIM group has two to six mannosyl units attached to the base with two acyl chains, which are either palmitic, stearic, or tuberculostearic acids. The most abundant PIM-type glycolipids are PIM2 and PIM6[43]. LM and LAM are extensions of PIM with either a mannose or arabinose branch. LM has α-(1→6)-linked mannosyl units branched out from the *O*-2 position of the mannose residue from PIM. LAM has LM with α-(1→5)-linked arabinosyl units[44, 45]. Therefore, there are three building blocks on LAM, including MPI core, polysaccharide backbone, and capping motif[46]. The core structure of LAM is

conserved in mycobacteria, whereas the capping motif moieties are different in different species. The three different capping moieties are identified to be mannose-capped LAM (ManLAM), phosphatidylinositol-capped LAM (PILAM), and non-capped LAM (AraLAM)[47-49].

The physiological functions of the glycolipids are not clear. Depending on their structure, they display different immune-modulating properties. ManLAM reduces the bactericidal activity of macrophages by inhibiting phagosomal maturation[50]. It is also able to load onto CD1 molecules for presenting to T-cells. The macrophage immune response to PILAM is different from ManLAM. ManLAM induces an anti-inflammatory response by negatively modulating the production of inflammatory cytokines, but PILAM elicits a pro-inflammatory response[51]. These LAMs are the markers for mycobacterial species. ManLAM is found in slow-growing mycobacteria such as *M. tuberculosis*, *M. leprae*, *M. avium*, and *M. bovis*, whereas PILAM and AraLAM are only found in fast-growing *M. smegmatis (M. smeg)* and *M. chelonae*, respectively[44].

1.5 Mycobacterial lipid-associated proteins

Based on the general functions of lipids described above, the physiological functions of lipids for pathogens such as mycobacteria can be classified into three types, which could lead to the research interest into these lipid-associated proteins. The main function of this group of proteins is to carry lipids for catalyzing metabolic reactions, for intracellular/intercellular signaling, or for lipid translocation due to lipid insolubility. Due to their insolubility in water, lipids are biomolecules that commonly require a

protein for transport in the cytoplasm of bacteria or through the cellular membrane. The type of lipid that binds to the protein is an indication of the function of the lipid-binding protein. The functions of the lipid-binding proteins that are associated with mycobacterial lipids can be categorized into three types: proteins involved in the lipid biosynthesis pathway, proteins that act as lipid transporters, and proteins that play a role in lipid recognition. The bacterial lipid-binding proteins involved in lipid biosynthesis and lipid transport are expressed and function intracellularly, whereas the proteins involved in lipid recognition are found in the extracellular environment. Due to the abundance of lipids in mycobacteria, lipid-binding proteins play important roles in maintaining cellular function. Approximately 30% of the genes are predicted to function in the pathway related to lipid transport or biosynthesis[42, 52].

The lipid-binding proteins involved in lipid biosynthesis pathways such as FASII biosynthesis and mycolic acid biosynthesis bind to substrates that have a long acyl chain. Acyltransferases bind to long fatty acids and transfer fatty acids to make phospholipids and triglycerides. These types of lipid-binding proteins have a catalytic function. Often, they are candidates for drug targeting because the functions of these proteins link them to an essential pathway that produces fatty acids as building blocks of lipids, or directly makes a final product as a component in the cell wall that stabilizes membrane integrity. The lipid-transport proteins serve as vehicles that carry lipids in the water-soluble environment, which is similar to the function of albumin in the blood plasma, and maintain a proper lipid distribution. The binding of lipid-transport proteins, unlike that of albumin, could be specific and assist to transport lipids that are synthesized in the

cytoplasm to their designated locations in the periplasm or though the membrane into the cell wall. Bacterial lipids are not only the main source of energy and components of the cellular structure, but also are used for sending signals inside and outside the bacteria. This expands the purpose of lipid binding from cellular metabolism to extracellular signaling, which is related to lipid recognition. There are proteins in host immune systems that recognize mycobacterial lipids and modulate the immune response, such as antigen-presenting proteins like major histocompatibility complex (MHC)-class and toll-like receptors such as Toll-like receptor 2 (TLR2). Phages also have lipid-binding proteins that construct the phage tail, which defines its host range by recognizing the mycobacterial glycolipids on the cell wall.

Lipid-associated proteins that are involved in fatty acid biosynthesis

Proteins involved in lipid biosynthesis often are essential for bacteria surviving in the environment because metabolic lipids are the fundamental elements for essential and physiological functions. One of the essential pathways in lipid biosynthesis is type II fatty acid biosynthesis (FAS-II), which is present in all bacteria sequenced so far, and makes precursors for phospholipids and mycolic acids. Typically, the enzymes catalyzing FAS-II have potential as therapeutic targets because the FAS-II system is only found in plants, bacteria, and some microorganisms[53]. The fact that FAS-II is absent in humans indicates that the enzymatic reactions related to FAS-II are attractive targets for designing inhibitors. In the bacterial FAS-II system, one discrete enzyme is in charge of the individual enzymatic reactions, unlike the multifunctional macro-enzyme

FAS-I found in humans[53]. Although the reactions of FAS-I and FAS-II are identical, the differences in the active sites provide an opportunity for designing potent inhibitors that are specific for the bacterial FAS-II system.

The genes and the chemical reactions related to FAS-II biosynthesis are well conserved in microbial genomes. Both Mtb and *Plasmodium* have FAS-II, and both have identical chemical reactions, but the enzymes of FAS-II are named differently in Mtb and *Plasmodium*. The names of the enzymes used hereafter are based on those for *Plasmodium*, which is related to the study in Chapter 4. FAS-II has a repeating cycle of four reactions, including condensation, reduction, dehydration, and reduction of carbon-carbon bonds (Fig 1.5). Each reaction is carried out by a different enzyme, called FabH (β-ketoacyl-ACP synthase III), FabG (β-ketoacyl-ACP reductase), FabZ (β-hydroxyacyl-ACP dehydrogenase), and FabI (enoyl-ACP reductase). The initiation of FAS-II starts by converting acetyl-CoA to malonyl-CoA, which is catalyzed by an enzyme complex of acetyl-CoA carboxylase (AccABCD). When FabD (malonyl-CoA:ACP transacylase) transfers malonyl to ACP (acyl-carrier protein), the ACP protein carries the intermediates to enter the four-step reaction cycle that is catalyzed by FabH, FabZ, FabG, and FabI[54, 55]. To date, the FAS-II enzymes related to the condensation (FabB, FabF, and FabH) and the carbon-bond reduction (FabI) have been identified as targets for antibiotic design in different organisms[55].

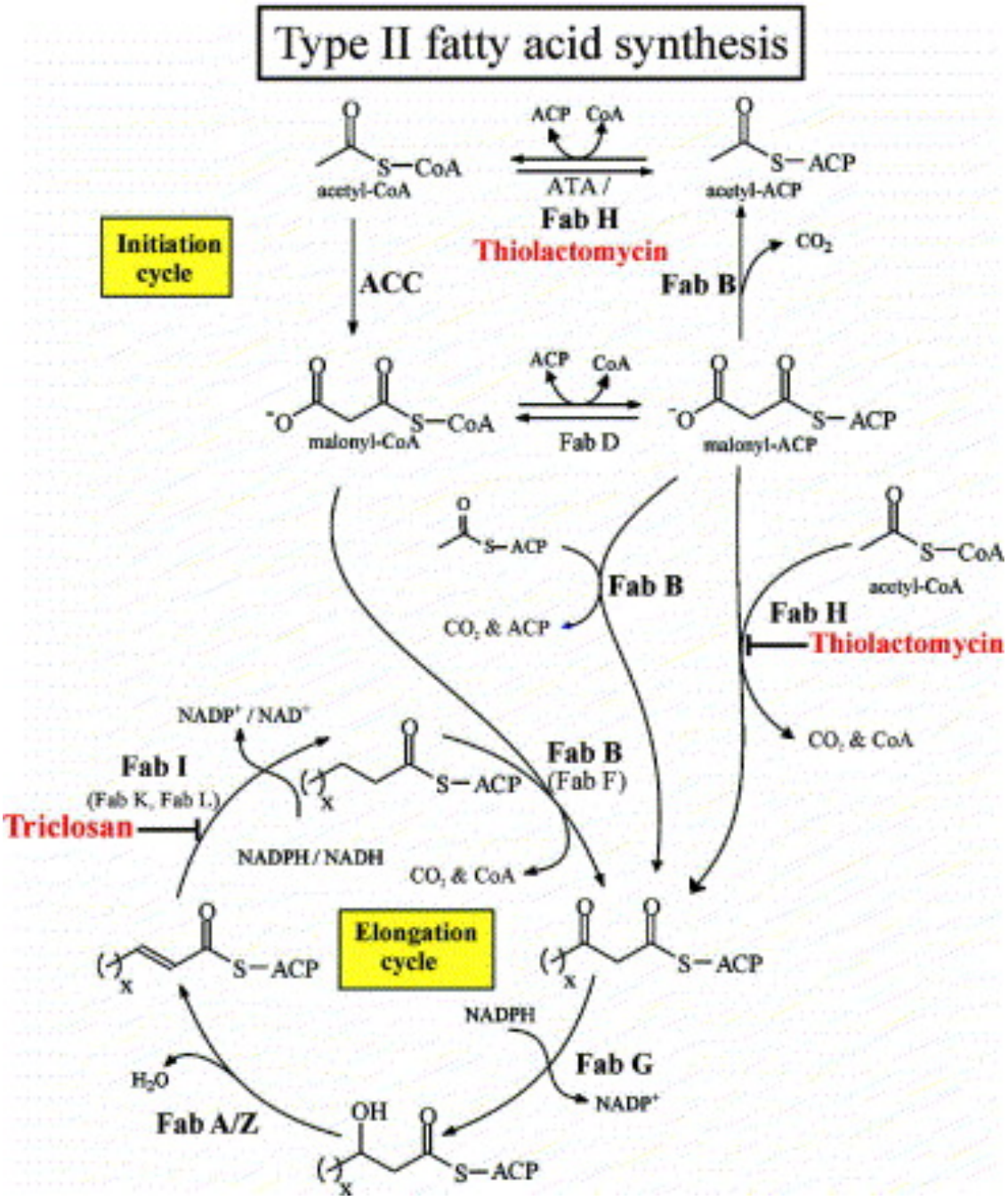


Figure 1.5 Diagram of fatty acid biosynthesis pathway[56]

Lipid-associated proteins that are involved in lipid transport

Lipid transport should facilitate the process of lipid biosynthesis due to the solubility issue. The reactions of lipid biosynthesis in bacteria occur initially in the cytosol, and the location of the final product is either in the plasma membrane or the cell wall. The translocation of metabolic intermediate lipids would require a lipid-transport system through the membrane. For example, LAM, one of the free glycolipids found in the cell wall, is the final product of the LAM biosynthesis pathway. The model for this pathway overlaps with the biosynthesis pathways for PI, PIM, and LM. The biosynthetic relationship is PI → PIM → LM → LAM[44]. The biosynthesis starts from the first step in PI synthesis and requires an inositol phosphate synthase (Rv0046) that catalyzes the conversion of glycose-6-phosphate to *myo*-inositol-1-phosphate [57](Fig 1.6). The dephosphorylation step is carried out by inositol-mono-phosphatase (IMP) to produce *myo*-inositol. The proteins Rv1604, Rv2701c, Rv2131c, and Rv3137 are homologs of IMP[58]. Subsequently, PgsA (Rv2612c) transfers diacylglycerol (DAG) from CDP-DAG to *myo*-inositol by esterifying a phosphate, and generates the phosphatidylinositol (PI) core[59]. Then, a mannopyranosyltransferase, PimA (Rv2610c), is the first enzyme to make PIM by adding mannose on PI at the *O*-2 position[60]. The next step involves Rv2611c, an acyltransferase, that acylates the 6-position of the newly added mannose and generates Ac1PIM1[61]. Further mannosylation is done by PimB (Rv0557c), α-D-mannose-α-(1→6)-phosphatidyl-*myo*-inositol mannopyranosyl transferase, adding mannose to the 6-OH of inositol and generating AcPIM2[61, 62]. PIMs contain 2–6 mannose attached to the PI core, and the α-(1→6)-mannopyranosyltransferases PimC

and PimD catalyze the formation of Ac1PIM3 and Ac1PIM4, respectively[63]. This is the branch point of the PIM and LM/LAM biosynthesis pathways[44]. Ac1PIM4 is the precursor for mannosylation at α-(1→2) by PimE (Rv1159c) to produce Ac1PIM6, or mannosylation at α-(1→6) by the MptA (Rv2174) / MptB (Rv1459c) / MptC (Rv2181) complex that feeds LM/LAM biosynthesis[45, 64]. MptB and MptA synthesize the α-(1→6)-mannan core. MptB catalyzes the proximal end of mannosylation by adding 12−15 mannose residues to the core, whereas MptA makes the α-(1→6)-mannan attachment at the distal end[65]. LM-containing branches of the α-(1→2)-mannan connected to the core are catalyzed by MptC[66].The transition from LM to LAM is the extension of 12−16 α-(1→5)-arabinosyl residues transferred by EmbC[67]. A few other proteins such as AftC, AftD, and AraT might have bi-functional roles in the arabinosylation of both LAM and AG biosynthesis in the cell wall[68-70]. This is a typical example, in that the proteins associated with the initial reactions are expressed in the cytosol and the final products are located in the cell wall. During glycolipid biosynthesis, there is a point where PIM is flipped from the cytosol to the plasma membrane. This could involve membrane proteins such as flippase to open a channel in the membrane and lipoproteins to facilitate the process. Mycobacterial lipoproteins are characterized into families based on the N-terminal signal peptide, which directs lipoproteins to the plasma membrane. The function of the lipoprotein LppX is involved in lipid transport on the PDIM biosynthesis pathway, in which the PDIM virulence factor is initially synthesized in the cytosol then translocated into the cell wall[71]. Only a few lipoproteins have an identified function that is similar to LppX. Therefore, the

glycolipids (PIM, LM, and LAM) also could require lipoproteins to assist lipid transport at a point between AcPIM4 and AcPIM6.

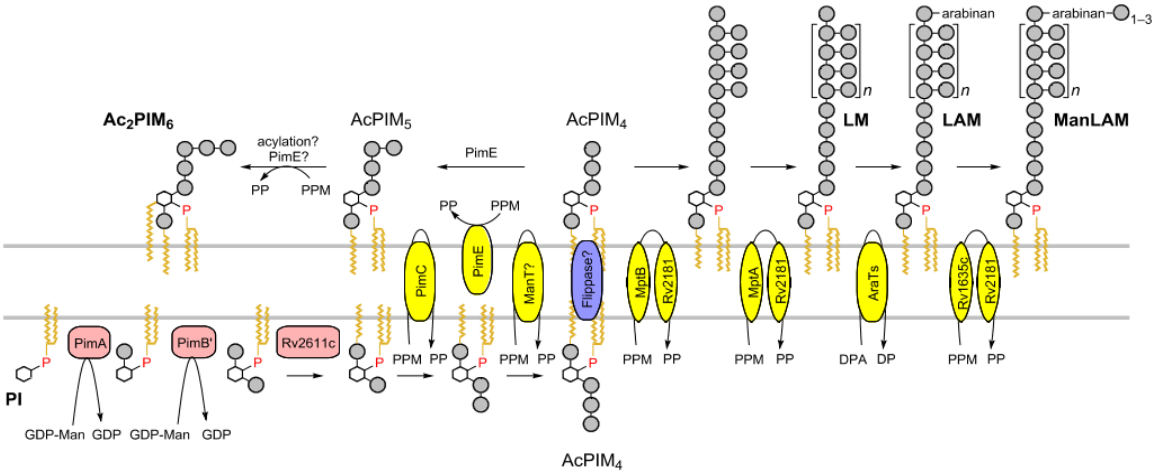


Figure 1.6 Lipoarabinomannans (LAM) biosynthesis pathway[72]

Host and phage lipid-binding proteins involved in recognition of mycobacteria

Glycolipids on the mycobacterial cell surface are located at important places, where proteins from other organisms can access them and use them for typing mycobacteria. Host and phage proteins that recognize mycobacterial glycolipids are this kind of lipid-associated proteins. In mycobacteriophages, they are tail fiber proteins (TFPs), which phages use to make initial contact with mycobacteria and to carry out infection specificity. In the host, they are pathogen recognition receptors that are used in the host immune system.

In phages, TFPs constitute the tip of a tail fiber, whose function is to make contact on the cell surface of bacteria. It has been demonstrated that the tip of TFP contributes to infection specificity[73]. Because mycobacteriophages only can cause infection in mycobacteria, their TFPs are likely responsible for binding onto the mycobacterial cell wall. Although the binding mechanism is still unknown, TFPs in mycobacteriophages have potential in selecting the specific glycolipids on the cellular surface that contribute to the recognition between phage and bacteria.

In the host immune system, mycobacterial glycolipids serve as biomarkers that pathogen-pattern receptors (such as TLR) and antigen-presenting molecules (such as CD1) can sense and detect the structural pattern to initiate the host defense system[74]. TLR proteins are evolutionarily conserved from worm to mammals. TLRs are associated primarily with innate immunity and are expressed on various immune cells, including macrophages, dendritic cells, B cells, and monocytes. A total of 12 TLRs are identified in mammals. They are type-I integral membrane glycoproteins, which are characterized by an extracellular domain and an intracellular domain. The extracellular domain of TLR that consists of 16–28 leucine-rich repeats is an immunoglobulin domain, and this domain also is the ligand-binding domain. The intracellular domain is a TIR domain, which is required for interaction with various adaptor molecules for activating the downstream signaling pathway[75]. To date, LAM from mycobacteria is recognized by TLR2[76]. The formation of TLR complexes such as TLR2/TLR1 and TLR6 responds to mycobacterial lipoproteins and tri- or di-acyl lipoproteins[77]. The antigen-presenting protein CD1 inspects through the endocytic compartment of the antigen-presenting cell

for either endogenous or exogenous lipid antigens. Subsequently, CD1 presents those lipid antigens to activate T-cells or NKT cells. Both PIM and ManLAM from mycobacteria are CD1-restricted lipid antigens[78].

1.6 Application of structural biology

Solving the protein structure of these lipid-associated proteins could provide information about protein function and insights into the mechanism related to protein function. When the essentiality of a gene is identified from genomic data, the structural study can fill in the gap between the genomic data and protein function. One of the major techniques in the determination of three-dimensional macromolecular structure is X-ray crystallography. Given the current technology, it can yield protein structures in a high-resolution range of 1–3 Å. Since Perutz and Kendrew solved the first protein structures of myoglobin and hemoglobin in 1953, there have been 89,212 protein structures deposited in the Protein Databank. Because of advances in the process of protein structure determination, such as high-throughput crystallization, liquid-handling robots, MAD-phasing method, and synchrontron source with automation, the rate of protein structure determination has increased rapidly in the past decade. Along with improvements in X-ray crystallography techniques, genome sequencing also is making progress and the genomic data of multiple species is available. Combined with the knowledge of the genome, cloning a gene is now precise and fast. Once the genomic DNA is obtained, it takes only a few days for protein production. However, protein crystallization and obtaining the phase are still the rate-limiting steps in this process of

protein structural determination. Furthermore, mapping the nucleotide sequences in the TB genome reveals that more than 50% of genes have no annotation and no identified functions[79]. When the genomic data become available, functions of the proteins get highlighted. Therefore, the focus of understanding biological mechanisms in living organisms has been switched to elucidate how proteins function. Consequently, the interest in protein function identification can be transitioned into a genome-wide project. For example, the Mtb structural and genomic consortium is putting effort into getting as many protein structures as possible and identifying their functions. The goal is to map all the protein functions through protein structural studies, which could be a useful database for therapeutic design of inhibitors.

The protein structural study can be useful in the function determination of target proteins and also in inhibitor optimization. Solving a protein structure provides insights into the different aspects of protein folding and conformation, which guide in the determination of function. When a protein function is unknown, the general protocol for annotating a function is to align the protein sequence to the database. If the alignment shows sequence identity higher than 30–40%, proteins with unknown function can be related to a protein family. A protein crystal structure shows the overall protein fold that may lead to the identification of a class of proteins, or could reveal a new fold for a new protein family. If this class of proteins has been identified, the function can be tested based on the structurally related protein family. For those proteins without known functions, a structural study is usually a guideline. When a protein structure shows a putative reaction site or cavity, an enzyme assay or a binding assay can be used to study

the protein function. In Section 2, the structural study of two lipoproteins of Mtb, LprG and LprA, is undertaken to perform the functional study. The key to understanding the function of both proteins is that the structures of both show a cavity, which indicates the presence of a binding / active site. Later, a binding assay is used to define the binding specificity. Mass spectrometry is a powerful tool for assisting functional studies. Therefore, the knowledge gained from functional and structural studies of proteins can expand the database of linkages from protein sequence to protein function. In addition, protein structure determination is a strategy used in drug design. Obtaining a protein structure forming a complex with an inhibitor is a probe to explore the potential interaction site that would help for inhibitor optimization. In Section 3, using the crystal structure of PfENR in complex with inhibitors in a structure-activity relationship study, the inhibition by a new series of inhibitors shows tight binding. Structural studies could be used as pilot studies of technology innovation. In Section 4, by identifying a tail fiber protein from a phage that has selective binding to mycobacteria, the protein structure becomes a guide for designing target therapeuticals at the initial stage.

2. STRUCTURAL STUDY OF LPRA AND LPRG*

2.1 Introduction

Proteins of various organisms undergo different post-translational modifications, such as lipid modification, for regulated functions. Lipid modification is a common process in bacterial organisms to make lipoproteins, which are the proteins secreted out of the membrane that function after post-translational modification[80]. The lipidation of lipoproteins facilitates protein anchoring to the cell membrane through the hydrophobic interaction of the lipidated acyl group and the cell membrane phospholipids[81]. In general, the bacterial genome contains 1–3% of genes encoded as lipoproteins[82]. This family of proteins has heterogeneous functions including nutrient uptake, signal transduction, adhesion, and conjugation, which are related to antibiotic resistance, transport, and extracytoplasmic folding of proteins. In addition to their functions, lipoproteins are associated with pathogen virulence such as colonization, invasion, evasion of host defense, and immunomodulation[80]. Therefore, mycobacterial lipoproteins are classified into a group of secreted proteins, and have been implicated in a broad spectrum of functions that participate in different pathways[83]. Either the

---

*Part of this section is reprinted with permission from "*Mycobacterium tuberculosis lipoprotein LprG (Rv1411c) binds triacylated glycolipid agonists of Toll-like receptor 2* " by Drage, M. G., Tsai, H. C., Pecora, N. D., Moody, D. B., Boom, W. H., Sacchettini, J. C. and Harding, C. V. et al. 2010. Nature Structural and Molecular Biology, **17**(9): p. 1088-95., Copyright 2010 by Nature America, Inc.

lipoproteins themselves or the pathways in which lipoproteins are involved are directly related to the survival of mycobacteria *in vivo* and their virulence[84, 85]. The interest in the study of lipoproteins has expanded from finding the physiological function to exploring the mechanism of virulence. There are extensive studies about lipoproteins that are unique to Mtb, including immunodiagnostics and the use of vaccine immunogens.

The lipoprotein biosynthesis pathway is a major virulence factor in Mtb[86-88]. The open reading frames of mycobacterial lipoproteins, which are known as pre-lipoproteins, are first translated to a full-length peptide chain. The N-terminal of these proteins contains a transmembrane domain, which is a signal peptide that directs protein localization and post-translational modification. This signal peptide is then cleaved after translocation of the lipoproteins across the membrane. In the lipoprotein biosynthesis pathway, these nascent pre-lipoproteins are processed by three integral-membrane proteins located in the plasma membrane: diacylglyceryl transferase (Lgt), signal peptidase (LspA), and N-acyl transferase (Lnt)[89]. These lipoproteins are often characterized by the presence of a conserved sequence, the lipobox, located in the C-terminal part of the signal peptide, and they consist of four amino acids (LVI/ASTVI/GAS/C)[83]. This lipobox motif functions as a recognition signal for lipid modification, which is made on the conserved and essential cysteine residue. The first step of lipoprotein biosynthesis is the formation of a thioether linkage catalyzed by Lgt, which attaches a diacylglycerol to the sulfur from the conserved cysteine in the lipobox. Then, LspA cleaves the N-terminal signal peptide by hydrolyzing the amide bond. N-

acyltransferase catalyzes the last step, which is the amino-acylation of the N-terminal cysteine residue[89]. After lipid modification, the mature lipoproteins are located in the outer leaflet of the cell membrane[89]. The mature lipoproteins are essential for Mtb to adapt to the environment, because disruption of the lipoprotein biosynthetic pathway in the lspA-deficient strain of Mtb resulted in decreased virulence and susceptibility to malachite-green[86].

When the mycobacteria invade the host, the receptors on the surface of the antigen-presenting cells (APCs) recognize the secreted proteins and the glycolipids from the mycobacterial surface[90]. These mycobacterial lipid and protein antigens induce the signaling pathways of the innate and adaptive immune responses[90]. In the innate immune response, the APCs (e.g., macrophages and dendritic cells) become phagocytes and engulf the Mtb to remove the bacteria from the host[91]. In the adaptive immune response, T-cells are activated by APCs to eliminate Mtb from causing further disease progression[92]. The Mtb lipoprotein pathway is crucial for T-cell activation, because the lspA-deficient Mtb shows loss of the T-cell response[84]. Therefore, mycobacterial lipoproteins have immunodominant properties of cell-wall-associated antigens. Despite having immunity as a protection from bacterial infection, bacterial-related infectious diseases like tuberculosis are still a major concern in public health. In humans, 10% of the population infected with Mtb proceeds with development of the disease[93]. One way Mtb survives the host defense system is by utilizing glycolipids (ManLAM) and lipoproteins to mediate the host immune system through receptors on the APCs, including Toll-like receptors and MHC-class proteins[32, 94]. Mtb with a deletion of

lprG fails to survive in mice and has decreased virulence in macrophages[95]. The LprG protein has the ability to induce TLR2 signaling through tri-acylation of the ligand Ac1PIM2[96]. This indicates that a small set of mycobacterial lipoproteins could contribute to mycobacterial virulence by suppressing the human immune response through the TLR2 signaling pathway. LprA is in a subset of the lipoprotein family that contains three other lipoproteins known as LprG, LppX, and LprF; it is able to induce the TLR2 signaling pathway[83, 97].

The mycobacterial lipoprotein LprA is a potent TLR2 agonist that induces the cytokine response and regulates APC function[97]. The TLR2 signaling induced by LprA is through a heterogeneous toll-like receptor complex of TLR2/TLR1 and TLR2/TLR6[90]. The acylation of the protein at the N-terminus is crucial for activation of TLR2 activity. It has previously been shown that triacylated lipopeptide (Pam3CSK4), a lipoprotein mimic containing six residues and a triacyl moiety at the N-terminus, is able to induce TLR2 activity. The structure of TLR2/TLR1 with Pam3CSK4 reveals how the acyl chains interact with the heterogeneous TLR complexes[98]. The truncated protein of recombinant LprA lacking the signal peptide results in a reduced signal through TLR2 when compared with that of the acylated LprA. To date, mycobacterial lipoproteins that have been characterized as TLR2 agonists (LprG, LpqH, and LprA) activate TLR2 activity by interacting with the TLR2/TLR1 complex[90]. However, LprA is able to induce the immune response through TLR2/TLR1 and TLR6/TLR2[90, 97]. LprG induces TLR2 signaling through one of the tri-acylated glycolipids (Ac1PIM2) that is carried by this lipoprotein, and more studies have shown

that mycobacterial lipids send a signal to mediate the host immune response. This suggests that the ligand of LprA, which will be defined by its structure, could assist in the elucidation of the role of this lipoprotein.

In this study, the crystal structure of LprA is solved and provides insight into a cavity, which is located at the center of the whole protein and has an entrance from the side. The crystallographic data shows that the cavity of LprA has a ligand bound when the protein is purified from *E. coli*. We identified that the ligand in the crystal structure is phosphatidylglycerol. The ligand identification is carried out in three bacterial species, including *E. coli*, *M. tuberculosis*, and *M. smeg*. Because LprA is a putative lipid-binding lipoprotein that is not involved in any catalytic reaction, ligand identification becomes one of the major indicators in finding the function of this lipoprotein. Utilizing the cavity of LprA to pull out the ligands from mycobacteria, the resulting ligand is analyzed by liquid chromatography-mass spectrometry (LC-MS). The differential LC-MS profiles in *E. coli* and Mtb indicate that a mixture of lipids bind to LprA, and most of the lipids share a backbone structure of phosphatidylinositol (PI). This finding is supported by the binding assay, in which PI shows a relatively strong affinity for LprA. LprA is on the same operon with two membrane proteins (Rv1272c and Rv1273c), which are homologs to a liposaccharide transporter protein MsbA from *E. coli*. Given our analysis on the ligand pool and the operon that contains LprA and MsbA in Mtb, the function of LprA can be predicted to be involved in the binding of phospholipids.

2.2 Results and discussion

Crystal structure of LprG

The data have been published and the paper is attached in Appendix A.

Crystal structure of LprA without signal peptide

LprA, which was cloned and expressed without a signal peptide, was purified with a His-tag on the N-terminus. Crystals of Mtb LprA grew from 25% PEG3350 and 0.1 M citric acid, pH 3.5, in the C2 space group with unit cell dimensions of a = 119.1 Å, b = 41.5 Å, c = 88.2 Å, β = 102.0°, and α = γ = 90.0°. The diffraction data was collected to 2.2-Å resolution. The structure was solved by using molecular replacement with an LprA homology model that was generated by the CPH program based on the LprG structure (PDB code: 3MH8). The initial model had four regions (Val69−Thr79, Thr98−Asp107, Asp120−Leu142, and Ans178−Pro201) that required model building to fit residues in the electron density map. After repetitive cycles of model building and refinement, the final model had a crystallographic R-factor of 0.21 and R-free of 0.25. There were two molecules of LprA present in one asymmetric unit, containing a 6-histidine tag, residues 45−235, two ethyleneglycols, one glycerol in one of the protein molecules in the asymmetrical unit (molecule B), and 309 water molecules. The C-terminal residues 236−244 could not be modeled due to the missing electron density. The buried surface between these two molecules was 2%. This indicates that the protein stays in solution as a monomer, as confirmed by analyzing on the gel filtration column. The two protein molecules in the asymmetrical unit have almost identical conformation,

with a root mean square deviation (RMSD) of 1.1 Å calculated with all observed α-carbons.

The overall structure of LprA has an alpha/beta fold, which is the same as the protein folding reported for the other two lipoproteins (LprG and LppX) from the family. This folding in LprA consists of four alpha helices and an anti-parallel beta sheet formed by 10 beta-strands (Fig 2.1a). A helix (α3) and a loop-helix (α2)-loop motif are at the center of the curved β-sheet. Two helices of α1 and α4 are parallel to α3, located on the same side of the β-sheet. There is a cavity formed between the helices and the beta-sheet. The cavity in this lipoprotein family shares two common features, a large volume and hydrophobicity, which also have been observed in the LprA structure from other organisms. This cavity, located at the curvature of the beta sheet, is solvent-accessible and the volume of the cavity is ~1300 $Å^3$. The depth of this cavity is 18 Å, and it is lined with the side chains of hydrophobic residues. In the X-ray data, there is an unknown electron density in the cavity, which is not from protein. The shape of this unknown electron density suggests that the bound molecules have two long chains.

The LprA cavity has an entrance that is surrounded by loops, including three short loops (Pro74–Arg77, Gly189–Ser190, and Val101–Glu103), residues Lys104–Glu106 on β4, and a long loop of the loop-helix-loop motif (Asp120–Asn138, connecting β5 to α2). Except for Arg77, Glu106, Ser135, and Ser139, the residues at the entrance are hydrophobic. There is a small opening constituted by Ser135, Tyr137, and Gly189, which connects the cavity entrance to a groove on the surface of the protein (Fig 2.1b). This groove, constituted by residues Thr186–Ala188 from α2 and residues

Tyr137–Gly148 from the centered loop-helix-loop motif, is at the conjoint region on the protein surface between α2 and α4.

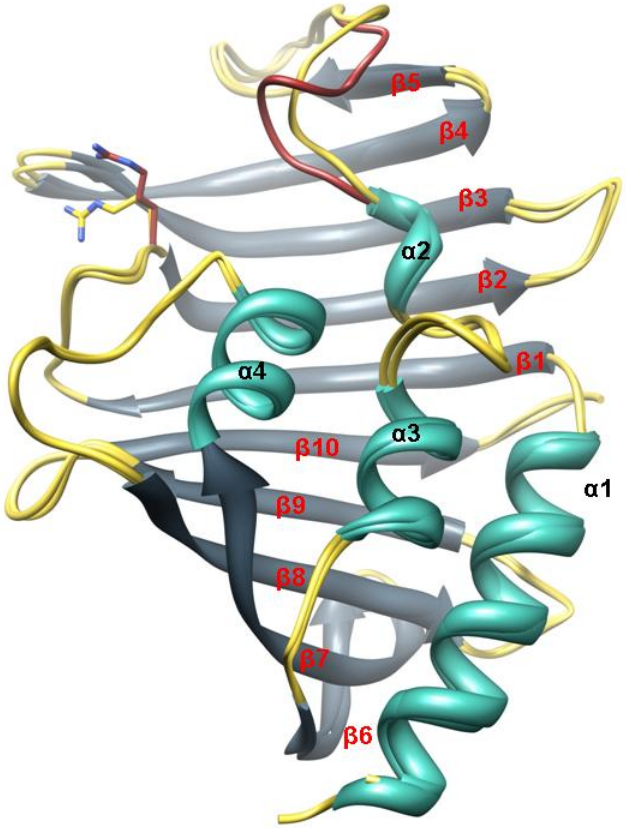


Figure 2.1a Superimposition of two crystal structures of LprA (molecules A and B) from an asymmetrical unit cell. The loops of molecule A are in yellow and the flexible loop of molecule B is highlighted in brown.

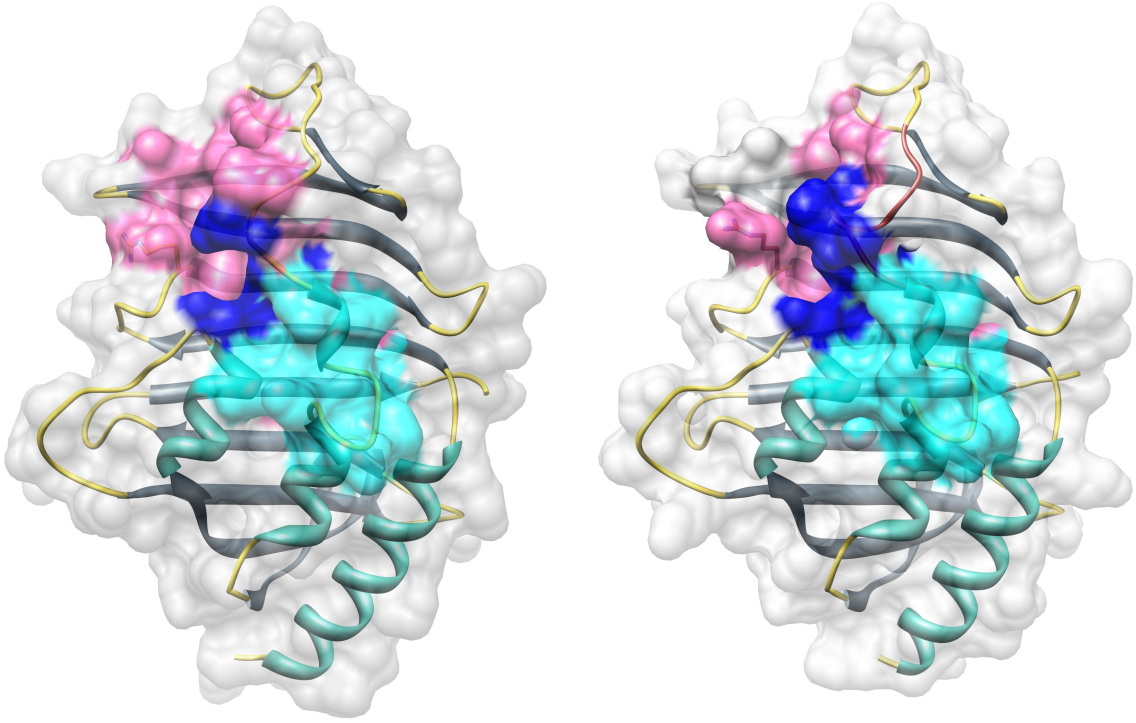


Figure 2.1b Overview of the LprA surface. The left figure shows the surface of one conformation found in molecule A of LprA. The right figure is the surface view made by the other conformation from molecule B of LprA. The hydrophobic groove on the surface and the cavity are highlighted with cyan and pink, respectively. The small portal connecting the cavity and the groove is in blue.

The overall structure of the two molecules (A and B) in the asymmetrical unit is highly similar, except for residue Arg77 and a long loop region (residues Asp120–Asn138) in the structure at the cavity entrance. The different conformations strongly suggest that this long loop on the centered loop-helix-loop motif, which covers half of the cavity entrance, is flexible. The carbonyl groups on the peptide backbone of Ile136 and Ser135 flip 147° and 108°, respectively, and cause a conformational change in residues Phe130–Asn138, which is a part of the long loop. The cavity was narrowed down by the γ-carbon of Ile136 shifting 5.8 Å and the ε-nitrogen of Arg77 flipping 101° (Fig 2.1a). Therefore, the cavity entrance of molecule B showed an opening size with dimensions of 7 Å × 15 Å, which is smaller than the cavity of molecule A (11 Å × 15 Å). The protein surface at Ile136 moves 4.8 Å to narrow the entrance, which is measured from carbon on the peptide backbone in molecule A to the carbonyl oxygen in molecule B. The position of Ile136 in molecule B blocks the small portal from the cavity to the groove. Apparently, the residues Phe130–Tyr137 of this flexible long loop region cause a major conformational change at the cavity entrance. Therefore, this flexible loop is likely to facilitate ligand binding and confer selectivity of the type of ligand.

The structure of Mtb LprA shows similarity to LprG and Lppx

The overall crystal structures of Mtb lipoproteins in this family (LprA, LprG, and LppX) are similar, and all of them have a large hydrophobic cavity. However, LprA and LprG are even more structurally alike because the cavities of LprA and LprG have the cavity entrances located in relatively the same position in each protein structure (Fig

2.1b). The superimposed structures of LprA and LprG show a high similarity with each other, with RMSD value of 1.7 Å calculated based on the α-carbon. Despite the overall resemblance of the protein folding, there are two regions that show different conformations, both of which are in the vicinity of the cavity entrance. The first difference is at the long loop region of the centered loop-helix-loop motif in LprA between α2 and β5 (Asp120−Asn138). This long loop region encompasses half of the cavity entrance and shows different conformations in the two molecules in the asymmetrical unit in the LprA crystal structure. The other region in LprA, which acts like the lid to the cavity, contains α4 (Ala181−Ala188) and a loop (Gly189−Lys198), connecting between α4 and β8. Here, the difference between the two proteins results in a smaller volume in the cavity of LprA. When superimposing the structures of LprA and LprG, the lid region shifts 4.3 Å as measured between the carbonyl oxygen of Ala188 from LprA and the carbonyl carbon of Ala153 from LprG. The position of α4 reduces the cavity size and narrows the cavity entrance in LprA. The smaller size of the cavity entrance and the cavity itself will constrain the ligand binding. Indeed, the structural data shows that the shape of the unknown electron density in the LprA cavity contains two acyl chains, whereas the LprG cavity can accommodate three acyl chains. This description implies that there is a possibility that LprA binds sphingomyelin and phospholipids (e.g., phosphatidylglycerol (PG), phophatidylcholine (PC), phosphatidylserine (PS), and phosphatidylethanolamine (PE)).

Ligand identification from the crystal structure by liquid chromatography and mass spectrometry

Considering the hydrophobicity of the cavity and the similarities in the architecture of the binding site to the cavity in Mtb LprG, the ligand of LprA can be predicted to be a lipid. To identify the bound ligand in the LprA crystal structure, it was extracted from the purified recombinant protein by treating with organic solvent (ethyl acetate) followed by strong acid denaturation of the protein. The extract was directly infused into the mass spectrometer (Bruker® MicrO TOF-Q-II). The results were analyzed by the software Compass DataAnalysis® (Bruker), which showed that the extract was a mixture of ligands that were presumably bound to the LprA protein expressed in *E. coli* (Fig 2.2a). The results also indicate that the ligands are tightly bound, because they are co-purified with LprA. The binding specificity was confirmed by comparing the mass spectra to the extract of Mtb LpqH, another Mtb lipoprotein that was expressed and purified in *E. coli* as a control. The ion peaks were found only in the Mtb LprA extract.

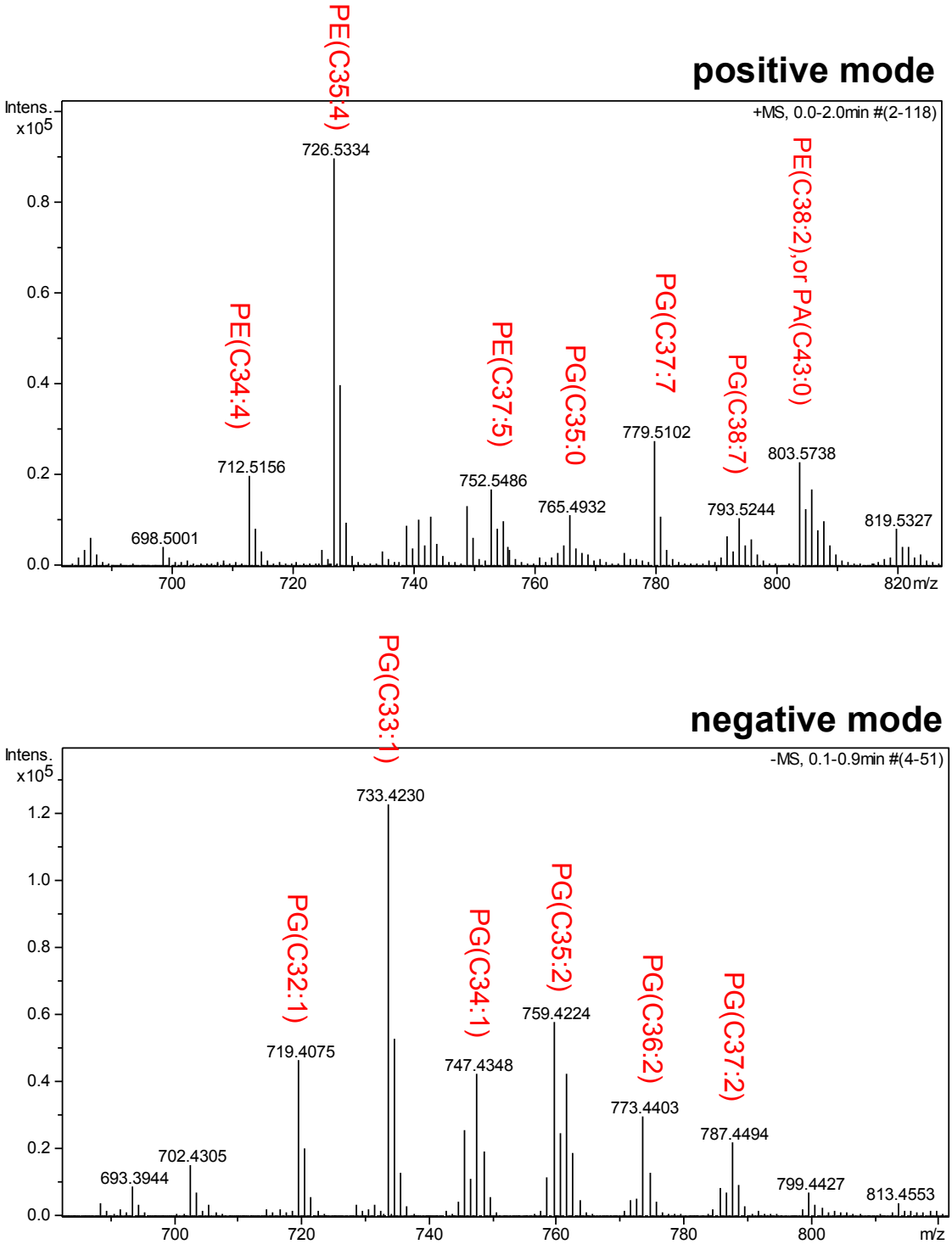


Figure 2.2a Mass spectral analysis of the ligand extraction from *E. coli* purified LprA. The extraction sample was injected into the mass spectrometer by the direct-infusion method and monitored in the ESI-positive and ESI-negative modes.

The high-resolution mass spectrometric data of the LprA extract showed one cluster of ion peaks in the ESI-negative and ESI-positive mode. The mass-to-charge ratio (*m*/*z*) of both ion clusters from the negative and positive modes are both in the range of 693 to 850. The molecular weight matched diacyl phospholipids, which also fit into the unknown electron density in the structural data. The ion peaks were preliminarily assigned by comparing them to the mass spectrometric database of lipids (Lipidomic Gateway). In the ESI-positive mode, two types of phospholipids were identified. The strongest ion peak (*m*/*z* 726.53) was phosphatidylethanolamine, PE(35:4), which has four unsaturated carbon-carbon bonds on the acyl chains. There were two more ion peaks (*m*/*z* 712.52 and 752.55), which also corresponded to the PE group, but they had different lengths of the acyl chains. The rest of the peaks in the ESI-positive mode were low intensity and corresponded to phosphatidylglycerol containing 32 to 37 carbons on the chain length (Fig 2.2a). In the negative mode, all of the major ion peaks were found to be in agreement with the molecular weight of PGs. The strongest peak (*m*/*z* 733.42) was PG(33:1). The rest of the ion peaks in this cluster was PG with 32 to 37 carbons on the diacyl chains. The data from both ionization modes indicate that LprA expressed in *E. coli* is mainly bound to PG and PE.

To quantify and further analyze individual components in the mixture, LC-MS analysis was performed. The ethyl acetate extract of LprA was injected into a liquid chromatography system (Agilent) and was monitored by high-resolution mass spectrometer (Bruker® MicrOTOF-Q II). The purpose of the liquid chromatography was to separate the ligand mixture based on polarity in a gradient of acetonitrile and water on

a reversed-phase column (C8, Waters®) at a flow rate of 0.5 mL/min. The extract was fractionated into three groups of eluents that were eluted with different percentages of acetonitrile (Fig 2.2b). The first group eluted in the range of 55−70% acetonitrile. It contained four components that were observed in both the positive and negative modes of ionization. Each of these four components in the first group represented one major ion peak, individually shown with a *m*/*z* of 501.4, 515.4, 529.4, and 543.4 in the positive mode that matched the molecular weight of LPE(18:1)+Na, LPE(19:1)+Na, LPE(20:1)+Na, and LPE(21:1)+Na, respectively. The negative mode also showed four components for the first group. The most intense ion peaks were *m*/*z* 297.1, 311.2, 325.2, and 339.2, representing LPA(15:1), LPA(16:1), LPA(17:1), and LPA(18:1), respectively. The ion peaks with *m*/*z* of 481.3, 495.3, 509.3, and 523.3 corresponding to LPG(16:1), LPG(17:1), LPG(C8:1), and LPG(19:1) also were observed, but with much lower peak intensity. The second group of eluents in the LC-MS profile eluted at 90% acetonitrile concentration in the solvent gradient, and was detected in both ionization modes. These ion peaks from the second group, showing *m*/*z* of 693.5, 719.5, 733.5, 747.5, 759.5, 761.5, and 773.5 in the negative mode, were attributed to PG(30:0), PG(32:1), PG(33:1), PG(34:1), PG(35:2), PG(35:1), and PG(36:2), respectively. The last group contained five eluents in the LC-MS profile that eluted at 100 % acetonitrile concentration. The ion peaks were assigned as fatty acids of FA(16:1), FA(17:1), FA(16:0), FA(18:1), and FA(19:1) with *m*/*z* values of 253.2, 267.2, 255.2, 281.2, and 295.3 in the negative mode, respectively. The LC-MS data showed that the ligands of

LprA were phospholipids (PG and PE), lysophospholipids, and fatty acids, which were known to be the most abundant lipids in the cellular membrane of *E. coli*.

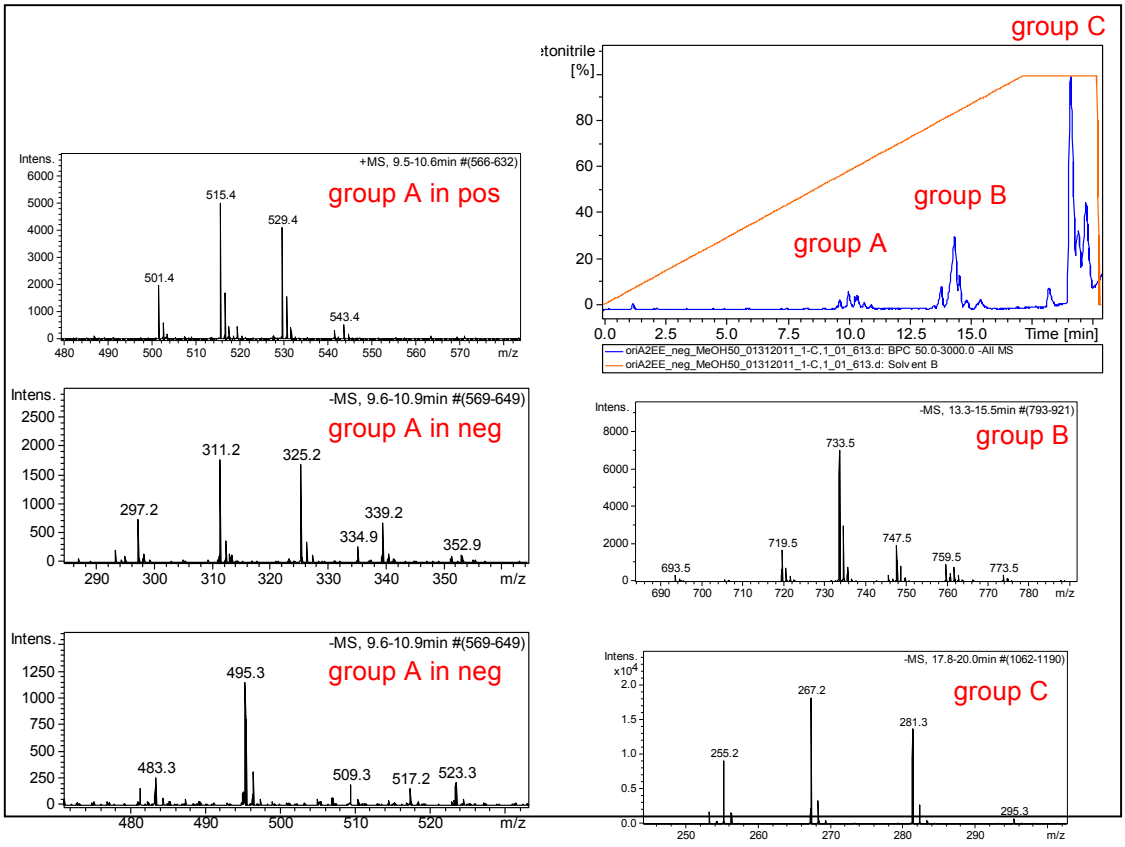
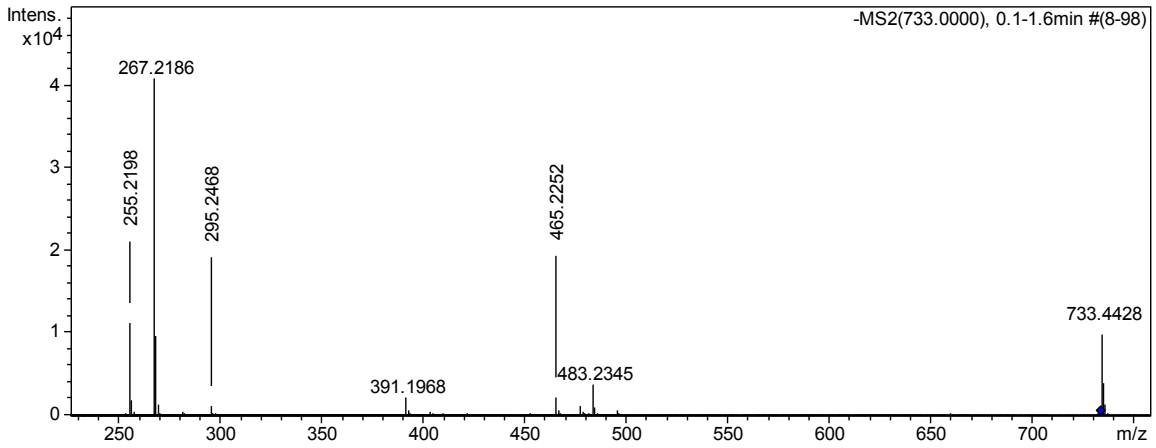


Figure 2.2b LC-MS profile of the *E. coli* purified LprA. This is monitored by the micrOTOF Q II in the positive and negative mode. The eluted peaks are integrated into three groups, based on the polarity. Except for the first group, the peak ions in both modes are the same. In the positive mode, lyso-PE was found and in the negative mode, lyso-PA and lyso-PG were identified. The second group of peaks is contributed by the diacylated PGs with various acyl chain lengths. Most of the ion peaks in the third group are fatty acids from FA(16:0) to FA(19:0).

Identifying the structure of the Mtb LprA bound ligands by the collision-induced dissociation method

The LC-MS profile indicated that the major molecules in the ligand pool matched the mass of PGs with various fatty acyl chain lengths, where the *m*/*z* was 719, 733, 747, 759, and 761. To confirm the fatty acyl chain lengths of the PGs, the collision-induced dissociation (CID) method was used to fragment the PGs. This was done by applying varying voltage to the collision cell inside the mass spectrometer that fragmented the PGs and generated free fatty acyl chains to be detected in the ESI-negative mode. The results of CID indicated that the molecule showing the peak at *m*/*z* 733.4 $[M-H]^-$ was indeed PG. This molecule was shown to contain two fatty acyl chains, FA(17:1) and FA(16:0), each with a mass of 267 and 255 in the collision mode (Fig 2.2c). Another molecule (*m*/*z* 747.5) that was identified as a PG had two fatty acyl chains of FA(16:0) and FA(18:1), showing a mass of 265.2 and 281.2. Using the same method to identify the rest of the molecules observed in the LC-MS profiling, the ion peak of the *m*/*z* 719 was attributed to PG(32:1), which was found to contain FA(16:0) and FA(16:1). Similarly, the other two ion peaks, which showed *m*/*z* of 759 and 761, were assigned to PG(35:2) and PG(35:1), with two acyl chains FA(17:1) and FA(18:1) for PG(35:2), and two acyl chains FA(16:0) and FA(19:1) for PG(35:1). The CID results supported our preliminary ion peak assignment by comparing with the lipidomic database, and confirmed that the majority of the ligands had two acyl chains, which matched what was observed in the structural data.



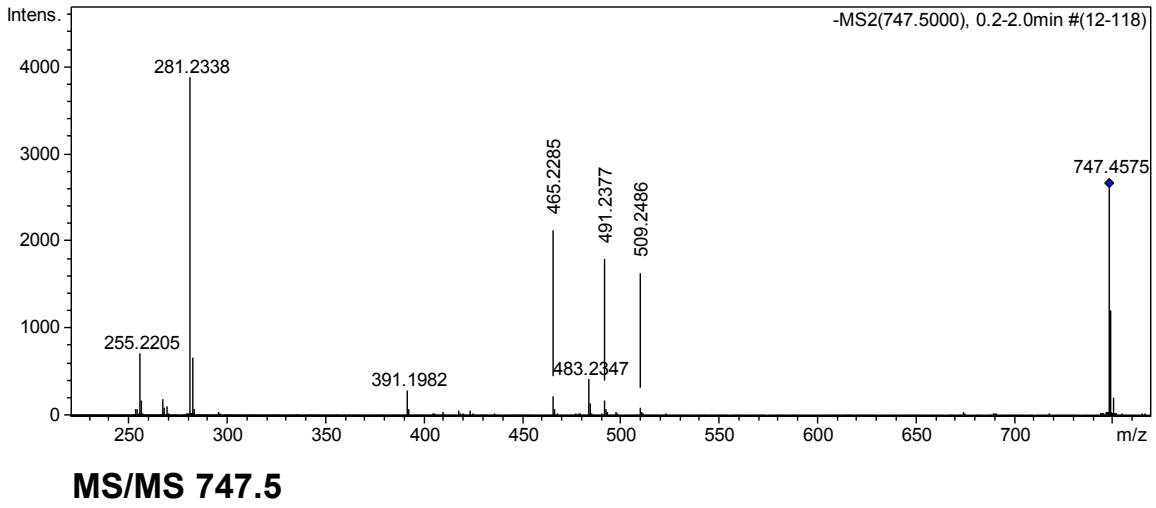


Figure 2.2c MS-MS of the bound ligands from the extraction. The potential ligands from the LC-MS result are analyzed by the collision-induced dissociation (CID) method. Varying voltages are applied to the collision cell, which results in fragmentation of the selected ions. These fragments are detected by mass spectrometry to help in identifying the molecular structure of the ligand. Two peak species of m/z=733.4 and 747.5 are selected to demonstrate the CID mass spectra.

The crystal structures of LprA with ligand bound in the cavity reveal two different conformations

Among all of the identified PGs in our study, PG(16:0,17:1) and PG(16:1,16:0) are the best fit in the electron density that is seen in the cavity of protein molecule A and B in an asymmetrical unit of LprA, respectively. Two acyl chains of both ligands are bound inside the cavity at identical positions in molecule A (Fig 2.3a) and molecule B (Fig 2.3b). Except for the length of one acyl chain, these two PGs have identical molecular structures. Despite the structural similarity of these two PGs, they reveal two different binding conformations, which are accommodated by the two different conformations of the protein shown in the two protein molecules in the asymmetrical unit.

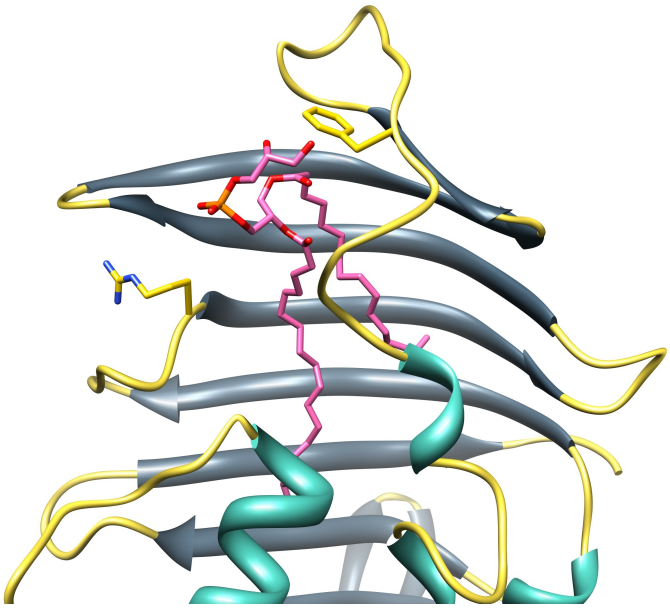


Figure 2.3a Crystal structure of LprA with PG(16:0, 17:1) in the cavity. (mol A)

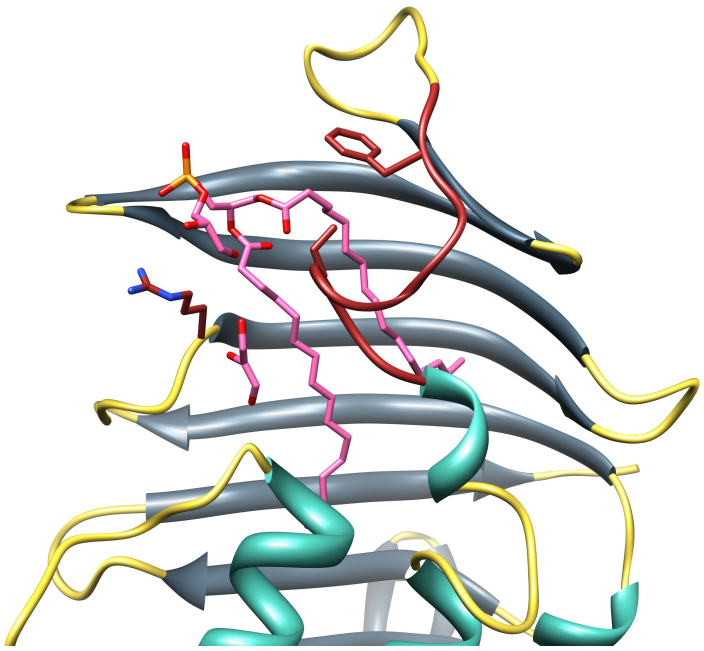


Figure 2.3b Crystal structure of LprA with PG(16:0, 16:1) in the cavity. (mol B).

The ligand PG(16:0, 17:1) in molecule A reveals a binding conformation that forms two hydrogen bonds to the residues on the long loop (Asp120–Asn138) connecting α2 and β5. The hydroxyl (-OH) on the *sn*-1 position of the glycerol head moiety forms a hydrogen bond to the carbonyl group of Phe130. The distance measured from the carbonyl of Phe130 to the *sn*-1-OH of the glycerol head is 3.2 Å. The other hydrogen bond is formed between the carbonyl moiety of FA(17:1) at the *sn*-2 position of the phosphoglycerate and the amine group of Ile136 at a distance 3 Å. A π-π stacking interaction between the side chain of Phe130 and the carbonyl group of FA(16:0) is observed in this conformation. However, these hydrogen bonds are not found in the structure of PG(16:0, 16:1) bound in molecule B. Instead, the *sn*-1-OH of the glycerol head turns away by 60° from Phe130, where phosphorus is set as center and the angle is

measured between the carbonyl oxygen of Phe130 and the oxygen of the *sn*-1-OH on the glycerol of the head group. Here, the phosphate moiety of PG(16:0, 16:1) forms relatively weak interactions with the side chain of Arg77 and the backbone of Gly102, where the distance between the oxygen on the phosphate moiety to the amine on the side chain of Arg77 and the amine on the Gly102 backbone are both 4.5 Å. The side chain of Arg77 flips 101° toward the phosphate moiety of PG(16:0, 16:1), which is measured between two ε-nitrogen of Arg77 from two conformations with a center fixed on γ-carbon. In this conformation, the phosphate moiety located outside the cavity makes contact with the residues Gln165 and Gln166 from the adjacent molecule formed by the crystal packing. The distance from the oxygen of the phosphate moiety to the amine group of Gln166 is 3.2 Å. Hence, the distance between the *sn*-1-OH of the phosphoglycerol and the carbonyl-oxygen of Phe130 is 9.6 Å, and the distance between the carbonyl group of *sn*-2-FA(16:1) and the amine of Ile136 is 8.1 Å.

The phosphate moiety of PG(16:0, 16:1) is bound at a different site from where the phosphate moiety of PG(16:0, 17:1) is bound. The superimposition of molecules A and B shows that the different ligand-binding conformations correlate to the long loop movement (Asp120–Asn138) and the rotation of the Arg77 side chain. The peptide backbones of Ile136 and Ser135 flip 147° and 108°, respectively, toward PG(16:0, 16:1), when the angle is measured by the carbonyl groups in the two conformations. Hence, the side chain of Ile136 also flips 142°. The angle is from the position of the β-carbon in molecule B to the α-carbon and β-carbon in molecule A. The β-carbon on the side chain

of Ile136 shifts 4.5 Å toward the ligand to form a hydrophobic interaction, and narrows the cavity entrance where the ligand PG(16:0, 16:1) binds.

Identifying the ligands of LprA protein from an Mtb extract

The Mtb LprA protein, similarly as other Mtb lipoproteins such as LprG and LppX, is expected to have a similar ligand-binding pattern. Mtb LprA overexpressed and purified from *E. coli* binds ligands specifically from the membrane of *E. coli*. Therefore, if LprA is mixed with the lysate from Mtb, the protein would be able to bind ligands that are found in Mtb. However, because the Mtb LprA was originally expressed in *E. coli*, the active site of the protein is already occupied by phospholipids from *E. coli*. Before the protein was mixed with an Mtb lysate, an apo-LprA was prepared by removing the bound ligands. This required two steps. The first step of removing ligand involved unfolding LprA protein in a buffer (20mM Tris pH 7.0 and 0.5 M NaCl) with 8 M urea to release the bound ligands, and the ligands were removed from the solution by adding 1 mM of β-cyclodextrin. Then, the protein was loaded onto a Ni-column and washed with 5 column volumes of buffer containing 8 M urea and 1 mM β-cyclodextrin. The protein was re-folded on the Ni-chelating affinity column by decreasing the concentration of urea from 8 M down to zero, followed by eluting in buffer with 300 mM imidazole. After dialysis with buffer (20mM Tris pH 7.0 and 0.2 M NaCl), LprA protein was homogenized with a cell pellet of Mtb and was lysed together with the cells. Then, the protein was purified again through a Ni-chelating affinity column.

The formation of the apo-LprA protein and re-loading of ligands from different extracts were confirmed by LC-MS. Three proteins samples, including the LprA purified from *E. coli*, the apo-LprA protein, and the protein mixed with an Mtb lysate, were extracted with the same protocol described above. Comparing the LC-MS profiles of these three extracts showed that ligands were removed before incubating LprA protein with the Mtb lysate, and that the LprA was loaded with ligand from the Mtb lysate. For example, the components that were identified in the LC-MS profiles of the *E. coli* purified LprA at 90–100% acetonitrile concentration had completely disappeared from the LC-MS profile of the apo-LprA protein. However, the extract of the LprA protein mixed with the Mtb lysate had two components in the LC-MS profiles that reappeared at the same percent of acetonitrile concentration. The mass (*m*/*z*) of these two eluents were 733 and 719, which matched the mass of PG(33:1) and PG(32:1), and these two ion peaks also were observed in the extract of the *E. coli* purified LprA. This indicated that LprA protein bound the ligands from the Mtb lysate (Fig 2.4a).

There is no LprA homolog in *E. coli*; therefore, LprA might have the capacity to load ligands that are found only in Mtb lysate. To identify the ligands bound to the LprA protein that were unique to the Mtb lysate, the LC-MS profiles of the extracts from LprA protein mixed with the Mtb lysate were compared to those of the extract from the *E. coli* purified LprA protein. To capture the ligand without knowing the polarity, two protein samples (10 nmol) were extracted with two organic solvents, methanol and ethyl acetate, and analyzed by LC-MS. The LC-MS profiles were analyzed by subtracting the reference run (the extract of *E. coli* purified LprA) from the sample run (the extract of

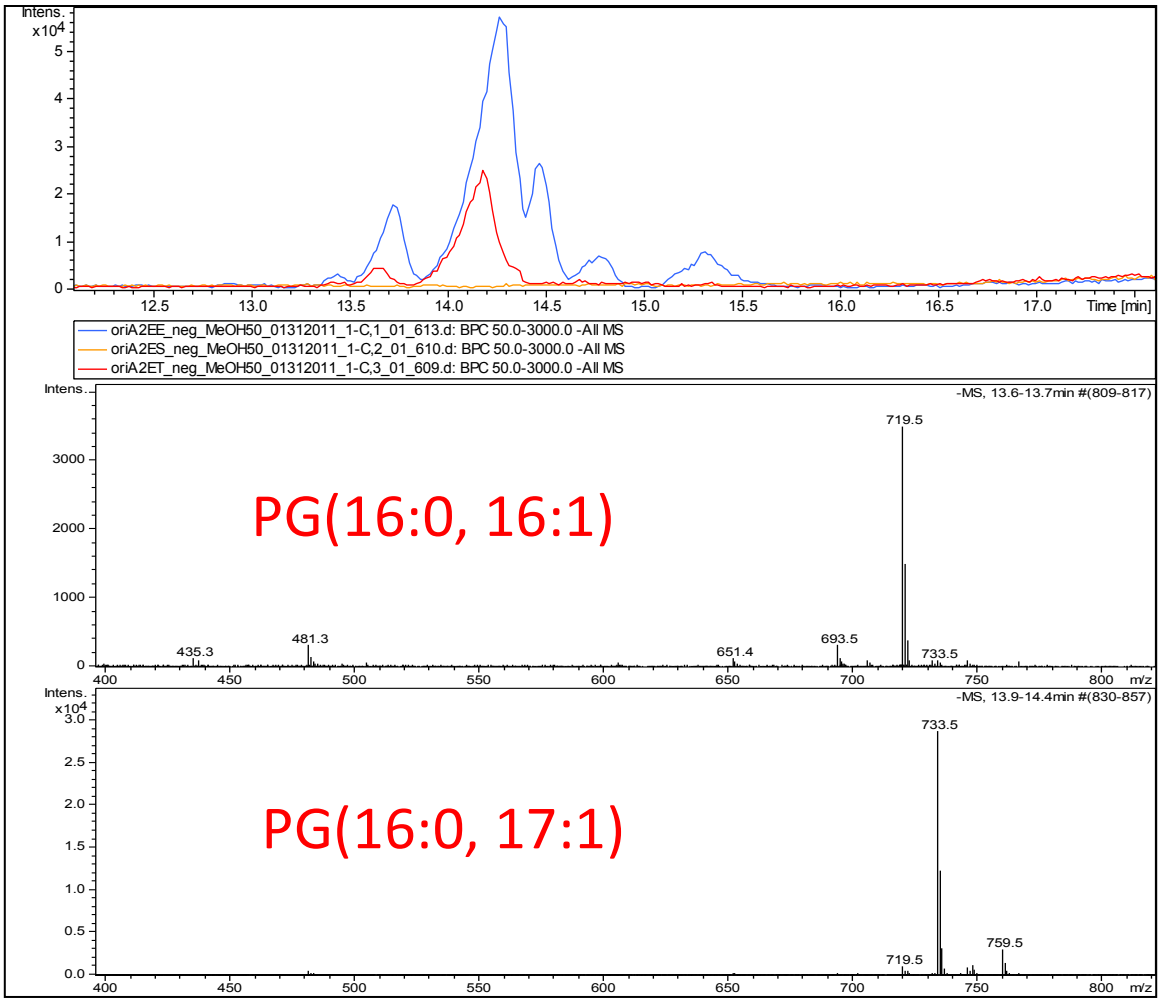


Figure 2.4a Identification of potential ligands from the Mtb lysate by LC-MS. The overlapped LC-MS profiles of the three extraction samples show the reloaded ligand from Mtb lysate. The LC-MS profiles of LprA containing ligands from *E. coli* (blue), ligands removed by cyclodextrin (orange), and ligands from Mtb (red) are shown in an overlapped format. Two peaks appear after incubation with Mtb lysate. These reloaded peaks represent *m*/*z* of 733.5 and 719.5, which match the exact mass of the best fit ligands in the crystallographic data.

LprA protein mixed with the Mtb lysate). Major differences were found between these two extracts in the ESI-positive mode. From the different profiles of the methanol extracts, the eluents were in the range of 45–60% acetonitrile concentration in the gradient, which indicated that the ligands extracted in methanol were more hydrophilic. There were two major eluents at the elution time of 3.5–5 minutes, and three more eluents that came later in the elution at about 7.5 minutes (Fig 2.4b). The first eluent in the LC-MS that was eluted at 46% acetonitrile contained lyso-PIM3(C17), which was observed in the MS at *m*/*z* 1094.58, 547.80, and 365.54, corresponding to the same molecule in three different charge states. The second major eluent was eluted at 59% acetonitrile, and the major ion peak could be attributed to phosphatidylinositol phosphate, PIP(38:5), which was observed at *m*/*z* 965.50 (*z* = +1) and 483.3 (*z* = +2). More than one species of PIP was found in the eluents, containing different acyl chain lengths that were 913.53 corresponding to PIP(34:3), 937.45 corresponding to PIP(36:5), 959.43 corresponding to PIP(36:5), and 987.48 corresponding to PIP(38:5). These PIP ion peaks also were observed in the ESI-negative mode. There were two ion peaks that were attributed to lysophosphatidylinositol with only one acyl chain, LPI(17:1) and LPI(20:4). Peaks representing PIM1 and PIM2 were observed at a lower intensity with *m*/*z* 1009.46 (PIM1[35:3]), 1038.50 (PIM1[37:1]), and 1170.64 (PIM2[35:3]). At later elution with organic solvent content close to 100% acetonitrile, there were three minor peaks, and each of them contained a single ion peak corresponding to LPG(16:1) of 483.26, LPE(17:1) of 488.25, and LPA[10:0] of 349.14, respectively.

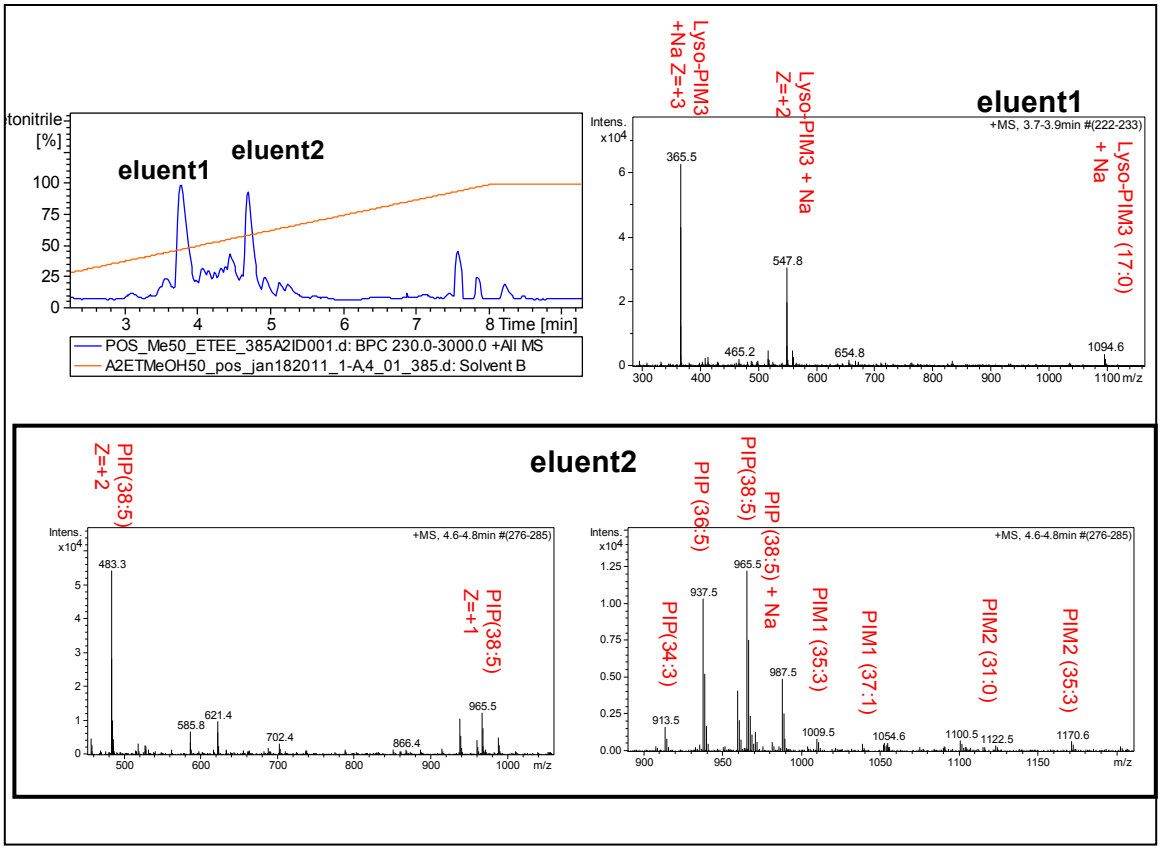


Figure 2.4b Differential LC-MS profiles of *E. coli* purified LprA and the Mtb lysate incubated LprA from methanol extraction. Two protein samples of LprA, one of which is purified from *E. coli* and the other that is incubated with the Mtb lysate, are prepared in the same amount (10 nmol) and extracted in methanol. Each sample is analyzed by the LC-MS method on a C8-reverse column. The profile shown above is the differential LC-MS profile, made by subtracting the profiles of these two samples.

Following the same experimental procedure in the previously described methanol extraction, the protein samples of the *E. coli* purified LprA and the Mtb lysate mixed LprA protein were extracted with ethyl acetate to extract any bound hydrophobic molecules. The difference in the elution profiles of the two samples showed four well-separated eluents, which eluted at the acetonitrile concentration of 87–100% (Fig 2.4c). The first two eluents contained small lipids such as monoglycerides (MG), lysophospholipids, mycerosic acids, and mycolipenic acids (MP). The mass spectra of the first eluent revealed that the *m*/*z* of the abundant molecules in that fraction were 359.30 [MG(C18)] and 431.32 [MP(C27)+Na] in the ESI-positive mode. The second eluent contained the same type of lipid molecules, but they had a slightly higher *m*/*z* 387.33 of MG(C20) and 459.36 of MP(C29)+Na. The third and fourth eluents had higher intensity shown in this difference profile, and each corresponded to the mass spectra signals 337.12 (*m*/*z*) and 365.15 (*m*/*z*), corresponding to lyso-PIP(16:1) and lyso-PIP(20:1), respectively. Both mass spectral ions of LPIP were detected in the +2 charge state. This was confirmed by observing the ion peaks corresponding to the singly charged state at 651.24 (*m*/*z*, $z = 1$) and 707.30 (*m*/*z*, $z = 1$) with a lower-intensity signal in the same mass spectrum.

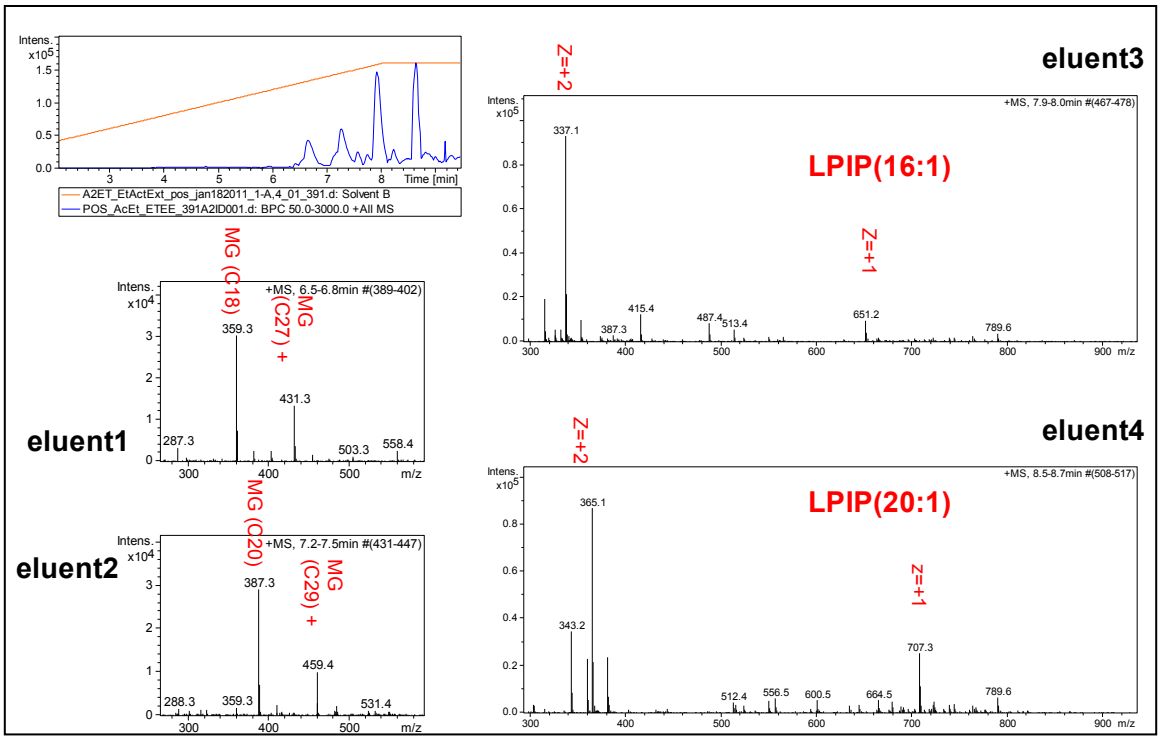


Figure 2.4c Differential LC-MS profile of the *E. coli* purified LprA and the Mtb lysate incubated LprA from ethyl acetate extraction. Two protein samples of LprA, one of which is purified from E. coli and the other that is incubated with the Mtb lysate, are prepared in the same amount (10 nmol) and extracted in ethyl acetate. Each sample is analyzed by the LC-MS method on a C8-reverse column. The profile shown above is the differential LC-MS profile, made by subtracting the profiles of these two samples.

LC-MS profile analysis of the extraction of the LprA expressed to contain the signal peptide in *M. smeg*

The differences in the LC-MS profiles of the LprA extracted from *E. coli* and Mtb indicated that a few new sets of lipids had been observed in the sample from Mtb lysate, such as mycerosic acids, mycolipenic acids (MP), phosphatidylinositol (PI), and

phosphatidylinositol phosphate (PIP). The abundance of PGs in Mtb was favored in binding to the LprA. PG was observed as the dominant molecule in the LC-MS profile, suggesting that the lipid content in the environment affects ligand binding to the LprA. This observation emphasized that the localization of LprA inside the bacteria is crucial for ligand identification. Therefore, it was hypothesized that extracting the ligand from the full-length LprA, where the N-terminal signal peptide would direct the LprA to the designated region, might yield the most physically relevant ligand.

The LprA contains a lipo-box at the end of the signal peptide. The signal peptide directs the protein to localize in the periplasm and anchor itself to the membrane. The lipo-box leads the LprA through the lipoprotein maturation process, in which the N-terminal signal peptide is cleaved and the signal peptide is swapped by a triacylated moiety attached to the N-terminal cysteine. Then, the LprA associates with the membrane through the triacylated moiety at the periplasmic side. Because the yield of the full-length LprA expressed in Mtb is extremely low, and *M. smeg* is shown to have a lipid composition closer to the Mtb, the full-length Mtb LprA protein obtained from the Clifford lab was expressed and purified from *M. smeg* (SmLprA). The software MetaboliteDetect (Bruker Daltonics) was used to process two sets of LC-MS profiles at the same time by comparing the signals of the sample to the blank. The software was used to subtract the background (blank control) from the MS profile of the sample, which resulted in magnification of the signal intensities from the ligand. The extraction of protein SmLprA followed the previously described procedure of acid denaturation and extraction in ethyl acetate. The extract of SmLprA was run on the LC-MS system with a

gradient of water to acetonitrile for 25 minutes, and was analyzed by the mass spectrometer (MicrOTOF-Q II, Bruker) in both the ESI-negative and ESI-positive modes. The LC-MS data showed a single major eluent in the ESI-negative mode. However, there were three major eluents in the ESI-positive mode (Fig 2.5). The first compound eluted at 30% acetonitrile, and mostly contained lysophosphates (LPAs). The last compound eluted at 88% acetonitrile and contained ion peaks with *m*/*z* ranging from 507.3 to 991.6 that could not be identified (each peak had an *m*/*z* interval of 44). The retention time of the second eluent in the ESI-positive mode matched the single eluent in the ESI-negative mode. Integrating the mass spectrometric peak from the negative mode showed that *m*/z 421.2 was the dominant ion. In the positive mode, the three ion peaks were at *m*/*z* 445.2, 440.3, and 423.2. Because an ion peak at *m*/*z* 844 was observed in the negative mode, the ion peak of *m*/*z* 421.2 in the negative mode was proposed to be the same ion species at a +2 charge state. The same observation in the positive mode confirmed that 432.2 and 445.2 both came from the same ion species showing *m*/*z* 421.2 in the negative mode. These ion peaks were all generated from a molecule with neutral molecular weight of 844, which matched the molecular weight of PI(35:4) from the lipidomic database. The other ion peak in the positive mode (*m*/*z* 440.3) followed a similar pattern, where it had a positive charge of +2 and a neutral molecular weight of 878, which also matched the molecular weight of a PI(37:1) in the database. Despite LprA bound to diverse forms of phospholipids, the LC-MS analysis of the extract of full-length SmLprA showed that PI was the major phospholipid bound to LprA when expressed in the designated periplasmic region.

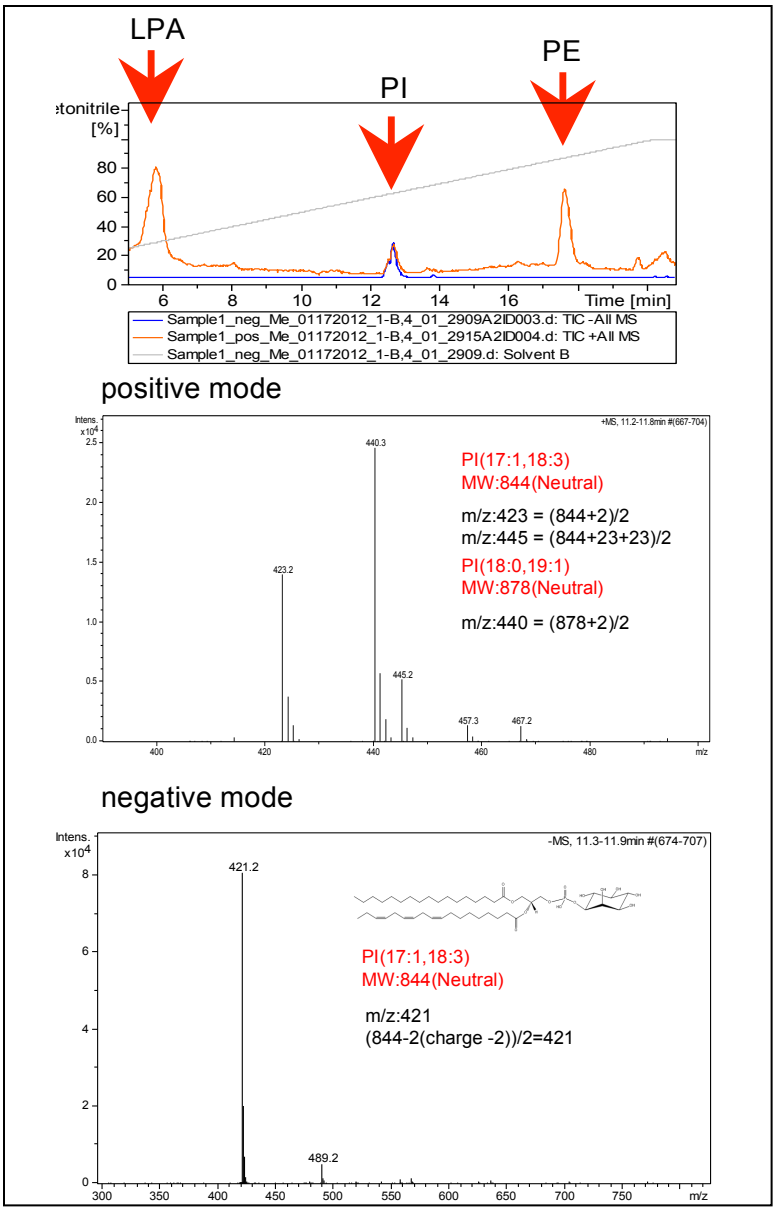


Figure 2.5 LC-MS profiles for analyzing the extraction of the full-length LprA purified from *M. smeg*. The top shows the LC-MS profiles of the LprA expressed and purified from *M. smeg*, shown in two ESI modes (the positive mode in orange and the negative mode in blue). The lower two figures illustrate the mass spectral data of the eluent (at the retention time of 12.5 min) in the ESI-positive and the ESI-negative modes.

The measurement of the dissociation constant of LprA and phospholipids in the intrinsic fluorescent-quenching assay

LC-MS analysis of three kinds of LprA protein samples (*E. coli* purified, Mtb ligand loaded, and *M. smeg* purified) showed that LprA is able to accommodate different types of phospholipids from the cellular membrane with various chain lengths such as PG, PE, and PI. However, the relative affinity of the lipids for LprA was not known. To determine the ligand affinity preference of LprA, a wide range of commercially available phospholipids and sphingolipids were selected to measure the dissociation constant (K_d) of LprA (Table 2.1). The affinity was measured by monitoring the change in the fluorescent emission at 340nm (the excitation is at 278nm) when LprA was titrated by phospholipid in buffer (20 mM Tris pH 7.0 and 300 mM NaCl). The fluorescent signal from residue Tyr137 located at the center of the cavity is quenched as phospholipids bind. The binding results in a decrease of the fluorescent signal, which was monitored by a fluorometer. The result showed that the tighter binding types of phospholipids are PE (K_d =0.35 mM) and PI (K_d =0.25 mM), which were detected in the previous LC-MS profiles. Cardiolipin (18:1) (K_d =0.36 mM) was shown to have a similar affinity for LprA. PS, which is another type of phospholipid that localizes in the bacterial membrane, had a fairly strong affinity for LprA, where both the K_d to PS(16:0, 18:1) and PS(18:1, 18:1) were in the range of 0.6–0.7 mM. By contrast, lipid molecules with a similar overall structure and similar chain length as the bacterial phospholipids but only found in mammalian cells, such as phosphatidylcholines (PC) and sphingolipids, showed

poor affinity for LprA. This indicates that LprA is selective for bacterial phospholipids, despite accommodating a broad range of ligands from phospholipids.

Surprisingly, among all the commercially available lipids tested, one of the major components of the bacterial membrane, PG, showed the lowest affinity to LprA. PG was shown in the LC-MS profile to be the dominant ligand of the *E. coli* purified LprA, and was observed to be taken up from a Mtb lysate. Therefore, it may be speculated that weak PG binding was very likely compensated for by its abundance in the bacterial membrane. After comparing the binding affinities of LprA to the different phospholipids, it was clear that the phospholipid head group played a critical role in defining the binding affinity to the protein. The fatty acyl chain length of the phospholipids does not significantly affect the binding affinity, even though the fatty acyl chains contribute to binding by forming van der Waals interactions in the binding cavity, as seen in the Mtb LprA crystal structures.

Table 2.1 Measurement of the dissociation constant.

Phospholipids	K_d (mM)
Cardiolipin (18:1, 18:1)	0.36±0.26
PG160181	2.6±1.1
PS160181	0.59±0.14
PS181181	0.71±0.62
PE160181	0.35±0.36
PC160160	1.3±0.7
PC170170	1.9±1.1
SPH_brain(18:0)	2.1±0.44
SPH_milk(23:0)	0.93±0.19
PI181181	0.25±0.04

Discussion

Based on the crystal structures showing the conformations of the bound PGs, two conformations reveal a similar binding pattern. The acyl chains of phospholipids interact with the side chains of the hydrophobic residues lined in the cavity, and the head group binds at the cavity entrance. Due to the identical core structure of phospholipids, one can infer that all of the phospholipids that are found in the extract bind to LprA in a similar binding pattern. Furthermore, based on the data presented in Table 2.1, the head groups of phospholipids affect the ligand affinity for LprA more than the acyl chains. Taking a closer look at the residues at the entrance to the cavity, it may be noted that the residues are mainly hydrophobic, except for a few residues of Arg77, Glu106, Ser135, and Ser190. These residues are the regions of the protein that are likely to interact with the phospholipid head group and act as the key factors for ligand recognition. The two sets of ligand-binding geometries show that the head moiety that connects to the phosphate group could form interactions with a hydrophilic residue (either Glu106 or Asp129). One of the geometries in molecule B shows that the phosphate moiety is positioned where the distance of the oxygen of phosphate to the ω-nitrogen of Arg77 is 4.4 Å. The interaction between residue Arg77 and the phosphate moiety of any phospholipid might be essential for tight ligand binding. In molecule A, the distance between the hydroxyl group of Ser135 and the oxygen of the phosphate group is 5.2 Å. This leads to a hydrogen bond formed from the carbonyl group of the *sn*-2-acyl chain to the backbone amide of Ile136. Among the ligands found in the extract, the PEs and PIs, which have a higher affinity for LprA, could be favored in both binding geometries because both binding geometries

result in the head group of the ligand potentially reaching to the hydrophilic residues of Glu106 or Asp129.

To get the LprA structure with the ligands identified from the extracts, the LprA protein mixed with the Mtb lysate and the *M. smeg* purified LprA was set up in the crystallization trial, but no crystals were produced. It was difficult to obtain the crystals of the *E. coli* purified LprA protein. Adding grinded seaweed powder as a nucleation agent was the only way to grow the crystals of the *E. coli* purified LprA protein, and obtaining crystals required high concentrations of the protein (70 mg/mL). The difficulty to produce protein crystals could be explained by the fact that the phosphate of PG in molecule B formed hydrogen bonds with the amide of Gln166 from an adjacent protein molecule in the crystal packing. Also, the *sn*-2-hydroxyl of the head glycerol moiety formed interactions to the residues Thr172 from the adjacent protein molecule through a water molecule. The distance of this water molecule to the *sn*-2-hydroxyl and the amide of Thr172 were 2.7 Å and 3.0 Å, respectively. Hence, PG binding might be crucial for growing the LprA crystals. Because ligands were found to be a mixture in the LC-MS analysis, it explained why LprA crystallized only at such high concentrations. There was identical core structure of phospholipids and a similar binding pattern in the cavity. Hence, the PG-bound structure can reveal how the head group of phospholipid interacts with the residues at the entrance. The amine group of the PEs could form hydrogen bonds to the side chain of Glu106 or Asp129, which may result in tighter binding when compared to the affinity and binding of PGs. The other major ligand group, PI, possibly also follows the same patterns of ligand-binding geometries. The head group of PIs, the

inositol moiety, interacting with either Glu106 or Asp129 could assist in increasing the affinity toward glycolipids (PIMs and PIs).

Structural comparison of the ligand binding to other Mtb lipoproteins, LprG and LppX

In our study, the crystal structure of LprA is shown to bind a phospholipid in the cavity. Similarly to LprA, the other two lipoproteins (LprG and LppX) in this lipoprotein family also have been reported to co-purify or co-crystallize with ligands bound in the hydrophobic cavity. These three proteins in this lipoprotein family all have a similar fold and a conserved lipid binding character. The crystal structure of LppX showed three long fatty acids (two C18:0 and one C22:0) in the cavity, which were co-purified from *E. coli* (Fig 2.1c). The crystal structure of LprG bound with a triacylated glycolipid (Ac2PIM1) was obtained in our previous study. Unlike LppX, in which the cavity entrance is located at the center of the helix-loop side of the protein molecule, LprA has more similarity to the LprG fold when comparing the relative location of the cavity to the overall structure (Fig 2.1a and 2.1c). The cavity shape in LprG and LprA is similar, and the ligands (Ac2PIM1 and PG, respectively), which contain an identical core structure of phosphatic acid, have a similar binding pattern in the cavity. The difference in the two lipoprotein-folds is in the volume of the cavity, so that LprA can only accommodate diacyl phospholipids. The residue of Ala188 and Gly189 on helix-4 (Ala181–Leu192) in LprA makes the cavity narrower by 4.3 Å, which prohibits a third acyl chain to bind (Fig 2.6).

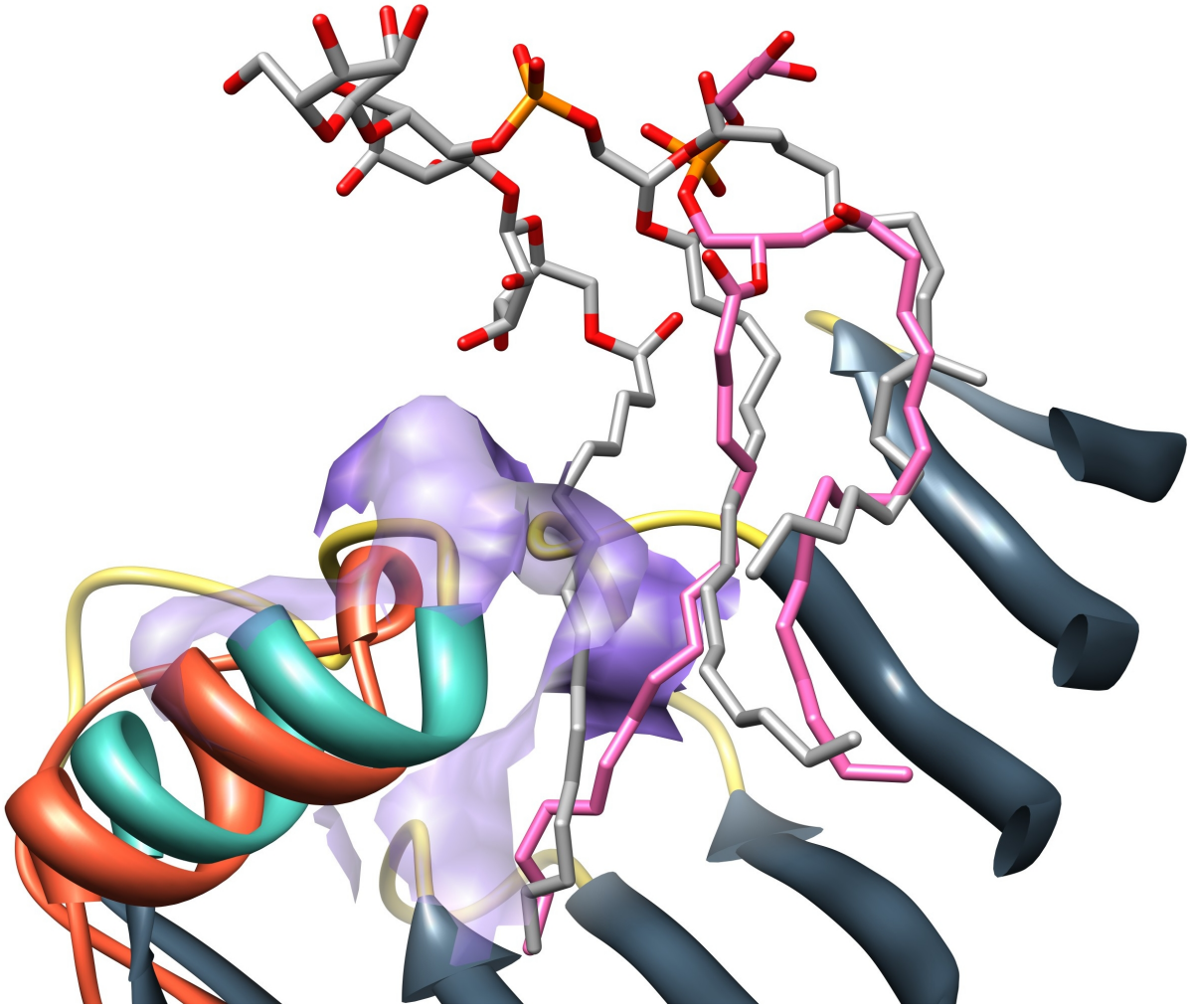


Figure 2.6 Structural comparison of the cavity in LprA and LprG shows a major difference on helix4. The structure of LprG (PDB code) with the ligand Ac1PIM2 (in white) is superimposed with LprA with PG (in pink). Ac1PIM2 has three acyl chains bound in the cavity of the LprG. The superimposed structures show that the cavity of LprA, unlike that in LprG, can only accommodate two acyl chains. The position of helix4 (in cyan) shifts close to the cavity to 3.6 Å, compared to that of the helix in LprG (in orange) The protein surface of that helix4 (shown in purple) completely blocks the space for the third acyl chain.

The result of the ligand extraction analysis highlighted a few major ligands that LprA bound from the mycobacterial lysate, and most of the ligand molecules identified in the extracts had a core structure of PI. From the Mtb lysate, LPIP, PIP, and PIM3 were the major ligands. The core structure shared by PIM3 and PIP has either three mannosides or a phosphate attached to the inositol moiety of the PI. PIM is a group of glycolipids that is found in the mycobacterial cell wall, and it has various numbers of branched mannose on the inositol moiety. PIP is produced when mycobacteria are in a stressed environment. The PIs are the dominant ligands in the extract of the full-length LprA expressed in *M. smeg*, which revealed that the LprA bound to specific ligands from mycobacteria.

We also observed PG and PE in the mycobacterial extract. This implies that the promiscuous ligand binding of LprA would have an influence in the protein function. The function of LprA could relate to the membrane proteins on the same operon. It has been known that LprG is in an operon with a MFS pump, and LprG is required for substrate transport mediated by the pump. Similarly, the gene *lppX* is located within a region in the genome that is involved in DIM biosynthesis (phthiocerol dimycocerosates, one of the mycobacterial virulence factors), and the function of LppX is related to DIM translocation through the cellular membrane. Because lprA is in an operon with two genes (Rv1272c and Rv1273c) that make a putative ABC transporter, the function of LprA (similarly to that of LprG) could be to mediate substrate transport. It is known that the cell membrane composition is dynamic and the PI turnover rate is fast. There is a possibility that LprA presents the cell wall environment to the cell for sensing and

adapting to stress. The neighboring genes, Rv1272c and Rc1273c, are the homologs to a lipid A/phospholipid flippase (MsbA) in *E. coli*, which transports phospholipids and lipopolysaccharides through membrane. This suggests that the promiscuous ligand binding is related to the function of LprA.

Lipid antigen comparison

Similar ligand-binding patterns are reported in all of the pattern recognition receptors (e.g., TLRs) and lipid antigen-presenting proteins (e.g., CD1b and CD1d). The lipid antigens bind to the antigen-presenting protein, with the acyl chains in the hydrophobic cavity and the head group at the cavity entrance. Mostly hydrophobic residues comprise the cavity entrance of the lipid antigen-presenting proteins, similarly as that in LprA, and there are two regions containing hydrophilic residues such as Arg, Asp, Glu, and Ser, which are conserved in the CD1 protein family. In the structure of the CD1 protein with ligand bound (PI-CD1b: 1GZQ, PIM-CD1d: 2GAZ, PC/PE-CD1b3: 3L9R, cardiolipin-CD1d: 3MA7), all the head groups of those phospholipids show interactions with these hydrophilic residues at the entrance. For example, the CD1d of PIM2 bound shows that the 3′-hydroxyl of the mannose interacts with the residue Arg79, and a distance of 3.3 Å is measured for the 3′-hydroxyl and the ω-nitrogen of Arg79. The oxygen of the phosphate moiety forms a hydrogen bond to the side chain of Ser76, which could relate to the interaction observed in LprA between the phosphate moiety of PG and Ser135. Therefore, the interactions between ligands and these hydrophilic residues at the entrance are definitely crucial for ligand binding and recognition.

Cardiolipin, which shows tight binding to LprA, is one of the lipid antigens that have been shown to bind to the cavity of the CD1b. The structural similarity between CD1b and LprA suggests that cardiolipin binds in the cavity of LprA, whereas the two acyl chains would bind inside the cavity of the LprA, similarly to that in PG. The other two acyl chains could to be exposed to the solvent and might reach the groove at the surface of the LprA. The hydrophobic residues at this groove could provide additional interactions for ligands such as cardiolipin [CD1b with ligand bound (PDB code: 3MA7)].

Most of the ligands bound to LprA are identified as the ligands for the antigen-presenting proteins, and the N-terminal lipopeptide of LprA also is the agonist of TLR2, which induces the cytokine response. There are two signaling pathways in which the mycobacterial lipids could modulate the host immune response, and LprA shows a potential ability to modulate the immune response. TLR2 forms a heterodimer with TLR1 or TLR6 and becomes a receptor. It can be induced to send a signal to the downstream pathway by binding a subset of the glycolipids on the di-acylated or the tri-acylated lipoproteins. LprA is the TLR2 agonist because it undergoes N-terminal triacylation, which would likely bind to the receptor complex of TLR2/1. The other complex (TLR2/6) is induced by the binding of the di-acylated lipopeptides, and it also is able to respond to the LprA. Therefore, one cannot eliminate the possibility that the diacylated lipids bound to LprA could activate the signal through the complex of TLR2/6. PIM1 and PIM2, which were found in the LprA ligand extraction, inhibit the LPS-induced inflammatory response through TLR4. The other antigen-presenting

pathway is related to the type I and type II CD proteins, in that both adapt to different sizes of lipid ligands to present to the T-cell. For example, mammalian PI(18:1/16:0) and mycobacterial PIM, which also were found in the LprA ligand extraction, are the antigens of CD1d, and are recognized by the T-cells or the NKT cells for antimicrobial host defense.

2.3 Methods

Cloning, expression, and purification of the truncated LprA

The gene *LprA* is encoded with an N-terminal signal peptide, which is for the lipoprotein translocation to the periplasmic region of the bacteria through the bacterial membrane. Overexpressing proteins that associate with the membrane might not have enough yield for the crystallization process. To increase the protein yield, the gene *lprA* was cloned in a truncated form without the signal peptide. The forward primer with an *Nde*I restriction site and the reverse primer with a *Hind*III restriction site were designed to amplify the target nucleotides from the genomic DNA of the Mtb H37Rv (BEI resource). This construct contained a His-tag and 190 residues (numbering 45–235) of the full length *LprA* (244 residues). The PCR product was digested with restriction enzyme and ligated into the pET30 vector (Novagen). The plasmid of pET30b-*lprA* with antibiotics selection (kanamycin) was isolated and transformed into the cell line *E. coli* BL21 (Novagen). The transformants were plated on Luria-Bertani (LB) agar medium containing 50 μg/mL kanamycin. A single colony was picked to inoculate the starter culture (20 mL). The cultures were grown to an OD_{600} of 0.6 at 37°C, and then the

starter culture was added to 2 L LB media. IPTG (1 mM) was added to the 2-L culture when the OD_{600} reached 0.6–0.8. After inducing protein production with IPTG, the cell cultures were placed in a shaker at 18°C for 16 hours. The cells were harvested by centrifugation at 3000×*g* for 40 minutes. The cell pellets were lysed by passing through a high-sheer fluid processor in buffer containing 50 mM Tris pH 7.5, 500 mM NaCl, 2 mM β-mercaptoethanol, and 10 µg/mL DNase. The lysate was centrifuged at 10,000×*g* for 60 minutes at 4°C to remove the cell debris from the soluble protein. The supernatant was passed through a 0.22 µm syringe filter before purification on the chromatographic system. The clarified supernatant was loaded onto a nickel column (GE Healthcare) in buffer 20 mM Tris pH 7.5, 500 mM NaCl, and 2 mM β-mercaptoethanol. After washing the Ni-column bound with immobilized LprA with 5 column-volumes of buffer, the protein was eluted in an elution gradient of the buffer (20 mM Tris pH 7.5, 500 mM NaCl, 2 mM β-mercaptoethanol, and 500 mM imidazole). The LprA protein was obtained at ~90% purity, concentrated to 100 mg/mL, and stored at −80°C.

Crystallization and structural determination

The truncated LprA (100 mg/mL) was screened under 576 conditions from five different crystallization screening blocks, which were set up in the sitting-drop vapor-diffusion method with the format of 96-well plates (Intelli-Plate, Art Robbins Instruments). The protein was pre-incubated with ground seaweed. Under each condition, the protein and the well solution were mixed in the protein well with the volume ratio of 1:1. After the crystallization plates were sealed for the equilibration, they

were placed on a shelf with limited vibration in a room with the constant temperature of 19°C. Crystals were obtained in the condition of 25% PEG3350 and 0.1 M citric acid pH 3.5.

The crystals were stable under the cryo-protectant (33% ethylene glycol and 67% well solution), and the flash-frozen crystals in liquid nitrogen were mounted on the home-source copper rotating X-ray generator (MMHF7000-RaxisIV++, Rigaku). The best crystal was diffracted to a resolution 2.0 Å, and the data were collected for 180° using a 1°-oscillation range per image. LprA was crystallized in the space group C2, indexed by HKL2000, with the unit cell dimension of a = 119.059 Å, b = 41.492 Å, c = 88.204, $\alpha = \gamma = 90°$, and $\beta = 102°$. The complete dataset was integrated and scaled using HKL2000.

The homology model of LprA was generated from the protein structure of LprG in the SwissModel server [99], because the protein sequence identity is 35%. This model was the input model for searching the initial phase information. The crystal structure of LprA was solved by the molecular replacement method. The initial solution was obtained by searching the input model at two steps (rotational and translational model searching) to determine the correct orientation of the input model in the unit cell. The initial LprA structure was generated in the MOLREP program from the CCP4 suite. The initial electron density was generated after a few cycles of the rigid-body coordinate refinement (Refmac, CCP4). The model building was done in COOT to fit the protein structure in the electron density map. The coordination of each residue from the model

building was refined against the X-ray data in PHENIX and Refmac. After rounds of iterative model building and refinement, the final model was obtained.

Ligand-reload

The truncated LprA was purified in *E. coli* and the crystal structure showed that the cavity was occupied with ligands from *E. coli*. To load the ligands from Mtb, the ligands from *E. coli* had to be removed first before introducing the metabolite pool from Mtb to LprA. Purified LprA was unfolded in the presence of urea at 8M. Then, beta-cyclodextrin was added at final concentration of 1mM. Unfolded LprA was loaded on the Ni-column and washed with 2 column-volumes of buffer containing 3 mM cyclodextrin (20 mM Tris pH 7.0 and 500 mM NaCl). When LprA was unfolded, the ligands from *E. coli* were released. An excess amount of cyclodextrin in the buffer could carry away those hydrophobic ligands during the washing step while LprA was immobilized on the Ni-column. The protein sample was extensively washed on the column without beta-cyclodextrin in buffer, and the concentration of urea was decreased in the buffer from 8 M to zero. During this washing step, LprA was folded back into an apo form without ligand occupying the cavity. Then, protein was eluted by the elution gradient of buffer, containing 20 mM Tris pH 7.0, 500 mM NaCl, and 300 mM imidazole. After dialysis in buffer without imidazole, apo-LprA was added to the cell suspension of Mtb mc2 7000. The cells and the protein were lysed together in the bead beater (BioSpec), and the cell lysate was centrifuged at 10,000×*g* for 60 minutes at 4°C

to remove the cell debris. The supernatant was purified on the Ni-column, following the same elution gradient protocol mentioned previously.

Ligand extraction

The truncated LprA was purified into three forms. Each of them presumably presented three different binding groups of ligands in the cavity. To identify the difference in ligand binding between these three samples, LprA was required to be removed first. Based on the crystal structure of the cavity with the hydrophobic residues lining up, the target ligands bound to LprA were predicted to be hydrophobic. Therefore, all proteins of these three forms were denatured in an acidic solution (1 M acetic acid) and were extracted with organic solvent (methanol or acetyl acetate) for further analysis. The purified full-length LprA was extracted with acetyl acetate or methanol, followed by protein denaturation in an acidic solution (1 M acetic acid).

LC-MS

Each sample obtained from the organic solvent extraction was transferred into methanol or a mixture of methanol/acetonitrile, which depends on the solubility in methanol. The analysis of the ligands from the extraction was based on the polarity by running a separation profile on the C8 column (Waters), which was connected to a liquid chromatography system (Agilent). The ligands were eluted in a gradient of solvent exchange from water to acetonitrile. The whole process of liquid chromatography was monitored by a mass spectrometer (Bruker Micro-QTOF). Data acquired from the

Micro-QTOF were integrated and analyzed by programs Compass DataAnalysis and MetaboliteDetect.

Binding assay

To determine the affinity of the potential ligands for LprA, intrinsic fluorescent spectroscopy was used to detect the signal decreasing as the ligand bound to protein. The LprA was unfolded to remove the ligand with beta-cyclodextrin, which yielded an apo form of LprA without ligand bound in the cavity. The wavelengths of excitation and emission were 280 nm and 340 nm, respectively, which were determined and optimized by scanning the protein at different wavelengths on a fluorometer (Cary Eclipse). Adding the stock solution of the phospholipids to the protein (3 μM) would quench the emission signal when the phospholipids bound to LprA. The quenched signal indicated the fraction of protein that was bound with ligand. Based on the plot of the concentration versus quenched signal, each data point was fitted in a linear equation to obtain the slope that represented the affinity to the LprA.

3. STRUCTURAL ANALYSIS OF PFENR FOR INHIBITOR OPTIMIZATION

3.1 Introduction

Malaria is transmitted by mosquitos and occurs mostly in tropical countries. The annual incident cases were estimated to be 219 million and caused 660,000 deaths in 2010 [100]. Regions with a prevalence of malaria are strongly and directly associated with poverty in those countries, and the high cost associated with preventative administration for controlling the spread of malaria makes the implementation difficult to sustain in poor and developing countries[101]. The total cost estimation for controlling and monitoring malaria globally according to the WHO malaria report is predicted to exceed USD 5.1 billion per year between 2011and 2020. However, the total funding was estimated to be USD 2.3 billion in 2011, which was considerably less than the amount needed to administer the disease control plan[101]. Given the favorable climate, weak infrastructure for implementation, and high cost, many countries are struggling to improve the incident and mortality rates. In areas with high incidence and transmission rates, malaria is the leading cause of illness and death. Malaria also is an opportunistic disease that affects populations with vulnerabilities in natural immunity such as young children who have not yet fully developed immunity and pregnant women who have weakened immunity[101].

Following the bite from an infected female *Anopheles* mosquito, the malaria parasites are transmitted into the human body and the incubation period takes from 7–30 days before the patient has the first-onset malarial symptom[102]. Usually, the most dangerous and fatal species, *Plasmodium falciparum*, has a shorter incubation period. Patients might have the three stages of symptoms from a cold stage (chills and shivering), hot stage (fever, headaches, and vomiting), and a final sweat stage (sweat, tiredness, and body aches), and are required to take the proper treatment according to the diagnosis[102]. The proper treatment might require more tests on the parasite species that caused infection, the clinical status of the patients, and the drug susceptibility. The WHO recommended a general treatment for patients who are suspected to have uncomplicated malaria caused by *Plasmodium falciparum* is most effective with artemisinin-based combination therapy[100]. Yet, in regions with lower incidence rates of complicated cases, the guideline is to take chloroquine treatment for 2 days when symptoms show up, as suggested by the CDC in the United States[102]. A high cure rate and low relapse rate are possible once patients comply with proper treatment plan. Despite the extensive efforts toward elimination of the disease, the resurgence of parasite strains with drug-resistance to effective drugs such as quinoline and antifolate (pyrimethanmine and sulphadoxine) catches our attention for a novel treatment design[103]. Ideally, the newly developed drug would work to expand the current repertoire of drugs and for use in combination therapies that could minimize the possibility of spreading drug-resistant strains.

The characteristic of having a plant-like plastid in *Plasmodium* provides new perspectives to search drug targets. *Plasmodium* belongs to a group of unicellular eukaryotes in the phylum of *Apicomplexa*, which is characterized by a non-photosynthetic plastid called the apicoplast. This organelle has four layers of membrane and is clearly a product of secondary endosymbiosis, in which the eukaryote has engulfed and retained another eukaryote with a plastid obtained by primary endosymbiosis from a prokaryote[104]. As an evolutionary homolog of plant and algal plastids, the apicoplast could be derived from an endosymbiotic *Cyanobacterium*.[105] Therefore, the prokayotic nature of the metabolic pathways is still contained in the apicoplast. In addition, a loss of apicoplast results in a block of replication and parasite death, indicating that apicoplast-related pathways are essential for parasite survival[106]. The finding of the apicoplast and its essentiality to the parasite are intriguing for the opportunities they present for therapeutic development. However, the genome of the apicoplast (encoding less than 50 proteins) provides limited information about its function except for basic metabolic processes such as DNA replication[107]. It was found that the majority of apicoplast proteins are nuclear-encoded. These peptides have an N-terminal bipartite leader for trafficking proteins into this organelle, which is associated with four pathways including fatty acids, isoprenoids, iron-sulphur cluster, and heme biosynthesis[105, 108]. These pathways are distinct from the pathways in mammals, and this highlights the potential for antimalarial drug targets.

Inhibition of the fatty acids biosynthesis pathway is a validated strategy as an appropriate target for developing antimalarial drug. *Plasmodium* has a FAS-II system,

commonly found in plants and bacteria. The four enzymes involved in this pathway are distinct from the type I FAS (FAS-I) pathway in mammals. In higher eukaryotes, the enzymes coordinating the FAS-I biosynthetic pathway are integrated into a multifunctional single polypeptide, and the substrate is carried by CoA[109]. By contrast, the FAS-II pathway utilizes acyl carrier protein (ACP) to carry the substrate for the iterative 2-carbon elongation, and employs four discrete enzymes to catalyze each reaction step, including β-ketoacyl-ACP synthase, β-ketoacyl-ACP reductase, β-hydorxyacyl-ACP dehydrogenase, and enoyl-ACP reductase[55, 110]. The last step is the rate-limiting step in the cycle of 2-carbon elongation, and it is catalyzed by enoyl-ACP reductase (PfENR). The mechanism of this reaction involves a hydride transfer from the cofactor NADH to the C3 position of 2-trans-enoyl of the substrate. The target study of PfENR has been done with triclosan, which is an effective antimicrobial agent.

Triclosan is a broad-spectrum antibacterial and antifungal agent, which was first characterized as targeting lipid synthesis in *E. coli[111]*. This finding was confirmed by comparing to drugs with known target pathways and the rescued enzyme activity at three individual mutated residues in *E. coli* FabI[111]. Researchers compared the percentage of acetate incorporated into lipids in the presence of triclosan and three other known antibiotics, including diazaborine (specific inhibitor of FabI), chloramphenicol (protein synthesis inhibitor), and ciprofloxacin (DNA synthesis inhibitor). They found that triclosan (0.24 μg/mL) inhibited 92% of acetate incorporation, which was similar to diazaborine that reduced acetate incorporation by 93%. By contrast, chloramphenicol and ciprofloxacin only reduced the incorporation by 19% and 2%, respectively.[111, 112]

Triclosan lost its potency of blocking lipid synthesis against the mutant FabIs. Triclosan and diazaborine have a similar minimal inhibitory concentration (MIC) value with three mutant strains[113]. The results again showed that triclosan targets lipid biosynthesis, especially on FabI, which is a protein homolog of PfENR. The Surolia group found that triclosan exhibited anti-malarial activity against the growth of *Plasmodium* with an MIC_{50} of 0.7 μ M, whereas cerulenin, a non-competitive inhibitor of fatty acid synthase, had an MIC_{50} of 20 μ M. When applying 2 μ M of triclosan in the media for measuring FAS-II activity, the amount of 1,2 [^{14}C]-acetate incorporated into fatty acids was shown to be reduced by 50%[114].

The essentiality of FAS-II to *Plasmodium* and the distinct active site of PfENR compared to the enoyl reductase from the mammalian FAS-I system make PfENR a qualified drug target for chemotherapeutic treatment. Combined with the potent inhibition of triclosan, the triclosan-bound PfENR structure serves as an initial survey of the chemical space in the active site. By substituting functional groups at the different positions of triclosan, the enzyme inhibition and the structural study of PfENR with triclosan analogs are correlated to generate a spatial analysis of the active site. This assists to explore the chemical space and the crucial residues for inhibitor binding. Therefore, the goal of structure-based inhibitor design of PfENR is to obtain inhibitors with higher affinity *in vitro*.

3.2 Results and discussion

Overall crystal structure of PfENR-inhibitor complex

The protein PfENR was purified based on the protocol published previously. The truncated PfENR was made with removal of the bipartite N-terminal signal sequence targeted to the apicoplast, and was constructed with an N-terminal His-tag for purification with the metal-affinity column. The pure protein was used with inhibitor during crystallization, and the crystals were produced from the condition containing 2.4 M ammonium sulfate and 0.1 M sodium acetate pH 5.6. The crystals with inhibitors or substrate were obtained by either soaking or co-crystallization. Inhibitor solubilized in acetonitrile was soaked in the PfENR crystals. This process took more than four weeks to get enough inhibitor occupancy in the active site. To efficiently obtain crystals of the PfENR protein in complex with inhibitors (days rather than weeks), co-crystallization was combined with a screening approach in a grid of 48 conditions, which was an optimized work platform to obtain crystals within two days. This grid was made with small variations in the ingredient concentration of the "hit" condition. Pure protein was incubated with inhibitors individually for one hour prior to setting up the crystallization plates. Then, crystals were mounted with the cryo-protectant paratone, and data were collected at the home-source X-ray generator with detectors (Raxis and Bruker CCD). The crystals from soaking and co-crystallization were isomorphous to the crystals of the PfENR-NAD+ complex, and were in the space group of P43212 with unit cells of (131.8 Å, 131.8 Å, 83.1 Å, 90.0°, 90.0°, 90.0°). Therefore, the crystal structures were all obtained by using the model of PfENR-NAD+ complex (PDB:1VRW) in rigid-body

refinement in REFMAC (CCP4), and refined with REFMAC (CCP4) or PHENIX to get the final structures with each inhibitor individually.

The structure of PfENR is reminiscent of the Rossmann fold and each subunit of PfENR is composed of 7 β-strands, which form a parallel β-sheet surrounded by 9 α-helices. The six beta-strands (β1−β6) from the parallel β-sheet compose two repeats of β-α-β-α-β in the topological order (β1-α1-β2-α2-β3 and β4-α5-β5-α6-β6). Between α7 and α8, there is a low-complexity insertion region, where the electron density is missing. The overall protein structure is similar to other ENRs when compared in the superimposition of PfENR with ENRs from *T. gondii*, *B. napus*, *M. tuberculosis*, and *E. coli*. The RMSD (root-mean-square deviation) of the Cα positions for PfENR with ENR are: 0.6 Å for *T. gondii*, 0.3 Å for *B. napus*, 0.75 Å for *M. tuberculosis*, and 0.78 Å for *E. coli*. The active site of PfENR is surrounded by α2, α3, α7, and the long loop between β4 and α5. Within the active site, the cofactor NADH binding site is located in the enzyme between α7 and the C-terminal end of the β-sheet. Although ENRs share similar folds, there is a major difference in a region corresponding to the conformation of α7 (residues 318−321) in the PfENR structure. In the *E. coli* and *B. napus* ENR structures, this region is a flexible loop, showing open and closed conformations of each individual active site. The protein structure of *B. napus* reveals that this loop region has an open conformation, showing that the region where triclosan binds is solvent accessible. This loop of *E. coli* ENR is in a closed conformation and is similar to the structure of PfENR. Based on the proximity to the active site, the residues on α7 contribute affinity for inhibitor binding and substrate recognition.

The NADH binding site

The substrate NADH is located primarily between α7 and α6, but it also is close to the C-terminal portion of the β-sheet and α7, and it is wrapped by several connecting loops (Fig 3.1). NADH binds to Trp131, Phe167, Ala169, Ser170, and Asn218 with the adenine ring moiety. The pyrophosphate moiety of NADH interacts with the loop region, connecting between α1 and β1, and the loop region immediately following α4. The nicotinamide ring binds deep within the active site and forms interactions with Tyr111, Leu265, Tyr267, Tyr277, Ala312, Gly313, Pro314, Leu315, and Ile369. The carboxamide moiety of NADH forms hydrogen bonds with Leu315. Tyr277, conserved among ENR, is the site where the reaction with the acyl thioester carbonyl group of the substrate occurs.

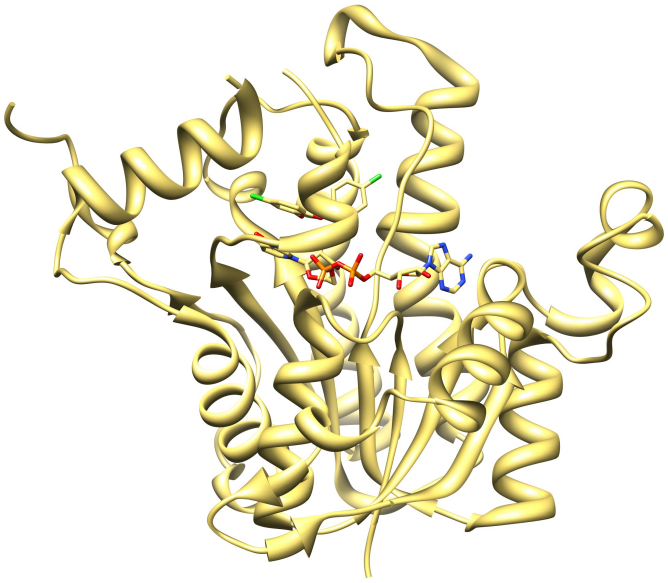


Figure 3.1 Overall PfENR structure.

Triclosan binding site

Triclosan, a potent inhibitor of PfENR, binds the substrate binding site by forming a non-covalent complex with NAD+ and protein primarily through hydrogen bonds (Fig 3.2). The crystal structure was reported previously, which was produced by incubating protein with triclosan and NAD^+, and showed that the binding orientation of triclosan is similar to the structures of the ENRs of *E. coli*, *M. tuberculosis*, *T. gondii*, and *B. napus*. In PfENR, triclosan binds between NAD^+ and α-helices α6 and α7 to form a stacking interaction with the nicotinamide ring of NAD^+. It also forms a hydrogen-bonding interaction with the 2′-hydrozyl moiety of nicotinamide ribose and Tyr277. In the PfENR:NAD^+:triclosan crystal structure, the π-π stacking interaction is formed between the nicotinamide ring and the phenol ring (Ring A) of triclosan (Fig 3.2). The 5-chloro moiety of this phenol ring, pointing toward the deepest portion of the active site in the closed conformation, formed van der Waals interactions with the side chains of Tyr267, Phe368, Ile369, Ala372, and Pro314. On the same ring, the phenolic hydroxyl is hydrogen-bonded to the 2′-hydroxyl moiety of the nicotinamide ribose and the oxygen of Tyr277. This residue is completely conserved in all known ENRs, and these interactions are crucial for the catalytic mechanism and inhibitor design. On the other ring, the dichloro-phenoxy ring (ring B) interacts with the pyrophosphate of NAD^+, and the distance of 2′-chloro from ring B and the oxygen from pyrophosphate is 3.5 Å. This ring B is located at the entrance of the pocket that is surrounded by the side chains of Asn218, Val222, Tyr277, Met281, and the backbone of residues of 217–231. The two

chloro-moieties, individually, interact with Asn218, Ala219, Val222, Met281(4-chloro), Ala217, and the oxygen of the pyrophosphate group from NAD+.

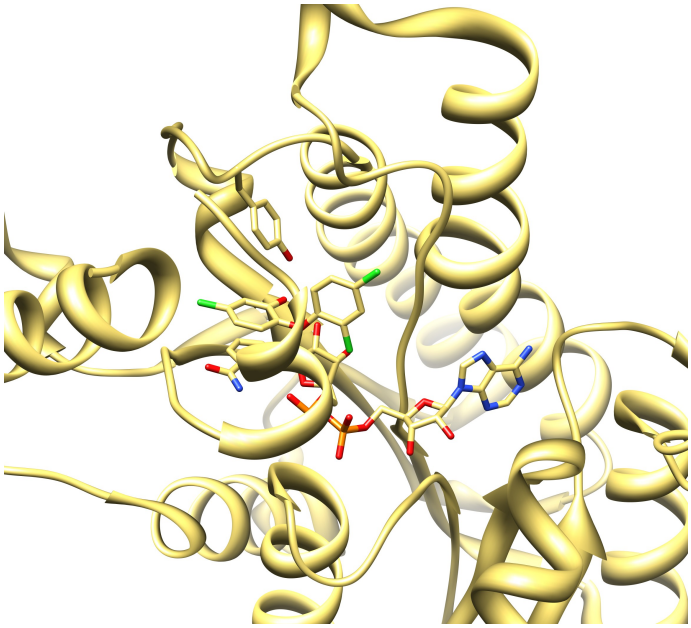


Figure 3.2 Active site of PfENR including triclosan-binding and NAD^+-binding sites.

Overall substrate binding in the PfERN-crotonoylCoA complex structure

Co-crystallizing PfENR with crotonoyl-CoA (C4:1), which is used in the enzyme assay, can explore where the real substrate, acyl carrier protein (ACP), would bind on the surface. This protein-NAD-crotonoyl-CoA complex was crystallized in the space group $P4_32_12$ and the crystal diffracted to 2.0 Å, showing the same unit cells (131.5 Å, 131.5 Å, 82.6 Å, 90.0°, 90.0°, 90.0°) obtained before. Therefore, the structure was solved by the same method using the Rigid Body program in REFMAC (CCP4 program

suite) due to the identical packing of protein molecules. Crotonoyl-CoA has four carbons linked to coenzyme A (CoA) through a thioester bond. In the structure, the butyl group on the crotonoyl-CoA binds at the triclosan binding site between Tyr267, Tyr277, and the nicotinamide moiety of NAD+, forming hydrophobic interactions with Tyr267 (5.0 Å) and Tyr277(3.8 Å) (Fig 3.3a). The carbonyl at the C4-position of the butyl group forms a hydrogen bond with the 2′-hydroxyl on the ribose of NAD+ at a distance of 3.2 Å. The butyl chain of the CoA pointing into the deepest part of the active site leaves the adenine-ribose-phosphate moiety outside and the pantothenate moiety at the entrance portal. This portal is constituted by α7 (residues 317–322) and a loop (residues 216–219). The thiol group on the beta-mercapto-ethylamine moiety of the CoA formed a sulfur-sulfur interaction with Met281 at a distance of 4.4 Å. The electron density for the pantothenate at the middle section of the CoA is not as clear as those of the butyl group and the adenine-ribosyl-phosphate group. This indicates the binding between PfENR and the pantothenate moiety is not tight, which allows crotonoyl-CoA to be flexible at the portal region. During crotonoyl-CoA binding, the side chain of Arg318 flips 108°, measured from the end nitrogen of Arg in the PfENR-triclosan structure to the same nitrogen atom in this structure with a center on γ-carbon of Arg. The conformational change allows the portal to become available to accommodate the pantothenate moiety binding at the portal between Val134 and Arg318. The adenine-ribosyl-phosphate group of CoA binds on the protein surface between α2 (residues 132–143) and the residues 185–189 on the loop connecting α3 and α4. At this region, which is far from the active site with a distance of 23 Å, an adenine group forms a hydrogen bond to Lys140 and the

bond length is 3.4 Å between the 3-nitrogen of the adenine and the ε-nitrogen of Lys140. The oxygen of the phosphate moiety on this phosphate ribosyl adenine group interacts with the nitrogen of Lys185 and the nitrogen of Arg186 and forms hydrogen bonds at a distance of 2.8 Å to Lys185 and 2.7 Å to Arg186.

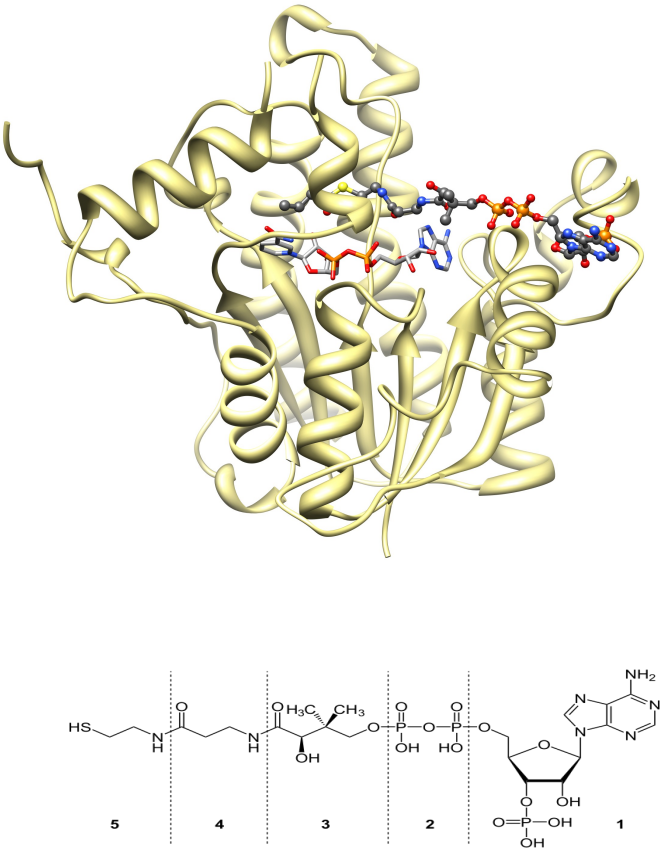


Figure 3.3a Overall view of crotonoyl-CoA binding in the active site (upper) and the structure of crotonoyl CoA (lower). The chemical moieties of crotonoyl CoA are labeled as following: **1**, 3'-Adenine ribose, **2**, diphosphate, **3**+**4**, pantothenic acid (pantothenate), **4**, beta-alanine, **5**, beta-mercapto-ethylamine, **3**+**4**+**5**, panthetheine

The PfENR structure with crotonoyl-CoA reveals two crucial regions on the surface for substrate binding (Fig 3.3b). One is solely contributed by Arg318 and the other is between Lys185, Arg186, and Lys140, where the adenine-ribosyl-phosphate group binds. The real substrate is carried by an acyl carrier protein (ACP), where one serine residue of ACP attaches to a phosphate-pantetheine moiety and the sulfhydryl group at the end of this moiety forms a thiol bond to the acyl chain. It is likely that ACP would bind through similar interactions as those observed in the crystal structure of PfENR-crotonoyl CoA. Indeed, the protein complex structure of BioH and ACP shows that ACP utilizes two sites to bind Arg residues on the surface of BioH(PDB: 4ETW). In the structure of ACP, one site is located at residues Asp38 and Asp41; the other is at Asp59 and Glu50 (PDB:3GZL). Therefore, these two major sites in ACP (PDB:3GZL) have potential for forming the interaction with PfENR when the enzymatic reaction occurs. As we proposed, when docking the ACP manually on the surface of PfENR, these two major interaction sites from ACP are able to bind to the two regions of Arg318 and Lys185/Arg186/Lys140 from PfENR (Fig 3.3C). Therefore, the pantetheine moiety from ACP likely binds between Val134 and Arg318, and ACP makes contacts to the surface of PfENR composed by α7 (residues 318−325) and α2 (residues 132−143). The residues Asp38 and Asp41 from ACP will be able to form interactions with Arg138 from PfENR, and residues Glu50 and Asp59 from ACP will be able to reach the residues of Lys185, Arg186 and Lys140 from PfENR (Fig 3.3d). Based on this model, ACP binding will not block the potential entrance for NADH. This reveals a tunnel for NADH to access the active site, which is constructed by Asp168, Lys240, Lys220, and Asn218.

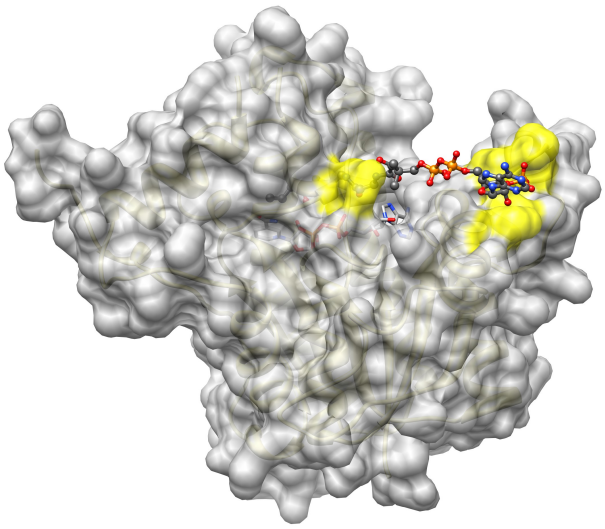


Figure 3.3b Surface view of PfENR with crotonoyl-CoA. The highlighted regions are the possible binding site for ACP-docking to form a complex.

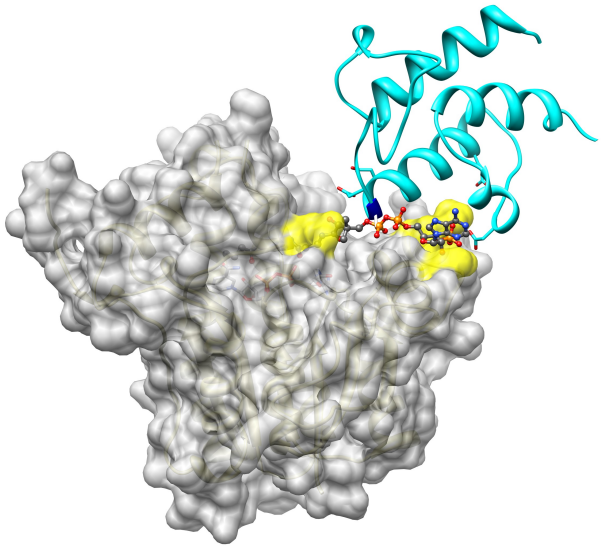


Figure 3.3c Model of ACP (PDB:4ETW) docking on PfENR

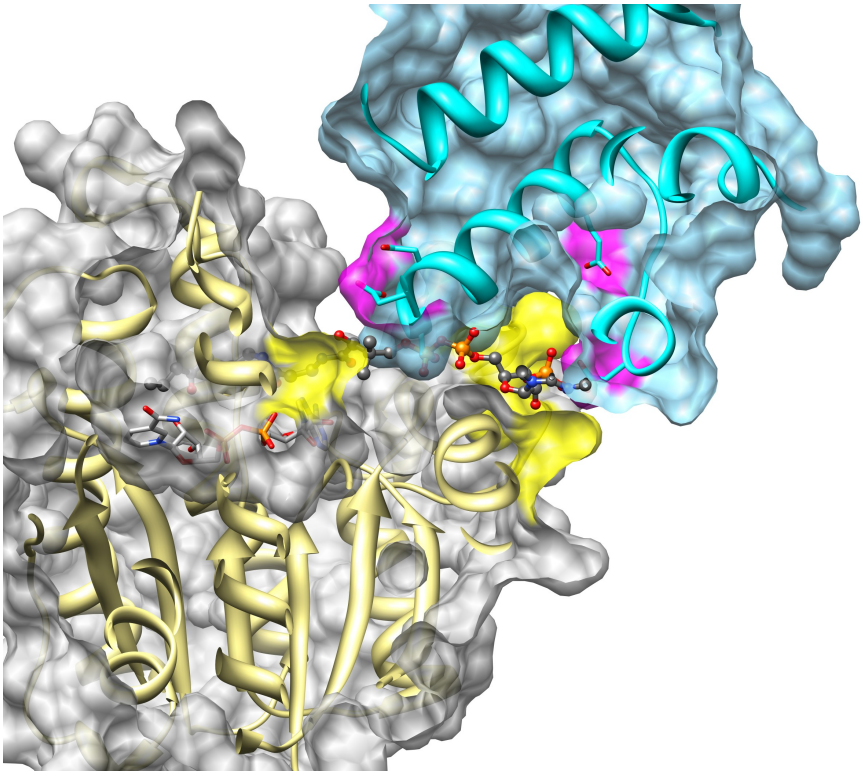


Figure 3.3d Possible residues making contact in PfENR and ACP.

Triclosan derivatives with modification at the 2-position

Inserted in Appendix B.

Triclosan derivatives with modification at the 5-position

Inserted in Appendix C.

Triclosan derivatives with modification at the 6-position

This is a continued structural study of exploring the PfENR active site to optimize the structure of triclosan, which has an enzyme inhibition (IC_{50}) of 70 nM. A

series of chemical modifications was designed to synthesize 4 clusters of triclosan analogs with functional groups attached at the 6-position (Table 3.1). The first cluster contains six compounds (**11**–**16**), and has small functional groups attached such as methyl and formyl group. The compounds in the second cluster (**21**–**25**) are slightly larger than those in cluster one, and have an amide backbone with different numbers of methyl moiety attached to the nitrogen. To expand the knowledge of the space in this direction of the 6-position, the compounds in the third cluster (aminosulfonyl group, **31**–**33**) and the forth cluster (ring group, **41**–**48**) have much bulkier functional groups attached. Compounds **45** and **46** in the fourth cluster are chlorinated at the different positions of the benzyl ring.

Table 3.1 IC_{50} of the 6-substitution triclosan analogs.

Compound	**R**	**IC_{50} (μM)**
11	COOH	25±18
1*2	CH_3	0.19± NA
13	CF_3	2.3±2.2
1*4	formyl	0.20±0.14
15	cyano	1.5±0.50
16	CH_2OH	4.4±1.8
21	$CH(OH)CH_3$	3.0±1.0
22	$C(O)CH_3$	4.7±1.6
2*3	$C(O)NH_2$	1.6±0.98
24	$C(O)NHCH_3$	5.6±3.5
25	$C(O)N(CH3)_2$	5.8±4.8
31	SO_2NH_2	4.2±2.1
3*2	SO_2NHCH_3	7.0±0.59
33	$SO_2N(CH3)_2$	6.0±3.2
41	2-propenyl	11±1.0
42	Tetrazol-5-yl	>10
4*3	Benzyl	0.19±NA
44	Phenyl	0.55±NA
45	m-chloro-benzyl	3.7±2.2
46	p-chloro-benzyl	3.9±3.0
47	CH(OH)-ph	3.4±1.4
48	C(O)-ph	1.6±0.50

All compounds were tested in the enzyme assay to measure the IC_{50} value. Compounds were diluted into a series of concentrations ranging from 100 μM to 0.01 μM and incubated with a mixture of 50 nM PfENR, 400 μM NADH, and 40 μM NAD+ for 30 minutes. The enzyme reaction was initiated by injecting crotonoyl-CoA (300 μM) and followed by monitoring the depletion of NADH by the omega plate reader. The result of the enzyme assay for the 6-position triclosan analogs showed that compounds with smaller substituted moieties and hydrophobic moieties have stronger affinity and inhibition against PfENR. Comparing the inhibition and the size of the modification at the 6-position among the clusters, the compounds in the first cluster with smaller modifications all have better inhibition with IC_{50} in the range of 0.2–5.0 μM, except for compound 11 with a carboxy moiety. Compound **12** (methyl) and **14** (formyl) show the lowest IC_{50}. However, substitutions of benzyl group, **43**, and phenyl group, **44**, also have low IC_{50} in the range of 0.2–0.5 μM. The correlation of inhibition versus chemical structure from the 6-postion modifications indicates that the additional hydrophobic interaction is more favorable to get potent inhibition.

Within these 6-position substituted derivatives, five crystal structures of PfENR:NAD^+ have been obtained with compounds **12**, **14**, **23**, **32**, and **43**. The binding conformations of these triclosan analogs elucidate an interesting result. Except for compound **32**, the other four compounds in this group bind to the active site with a completely opposite orientation that has not been reported previously (Fig 3.4a-d). In this new orientation, the positions of ring A and ring B are flipped compared to the triclosan-bound structure. Hence, unlike all triclosan analogs, the critical hydrogen bond

for tight inhibitor binding between Tyr277 and the phenolic hydroxyl on ring A is not seen in the structure of these four 6-position modified analogs. Instead, this hydrogen bond is replaced by an interaction between the 2'-chloro moiety from ring B and the hydroxyl of Tyr277, given a distance of 2.4 Å from the 2'-chloro to the oxygen of the Tyr277 hydroxyl. The two rings flipping results in the 6-position modifications reaching where originally the 3'-position modifications would point in the structure of triclosan. Now, the modified groups point outward from the active site to a new site, which is the NADH access tunnel. At this site, the protein surface for making interactions with the modified functional groups are through the side chains of residues Ala319, Ala322, Asn218, and Trp131, and the backbone of Ala217 (Fig 3.4e). This supports the result of the enzyme assay, in that the hydrophobic moiety here contributes to the inhibitor binding and the bulky modifications are not in favor.

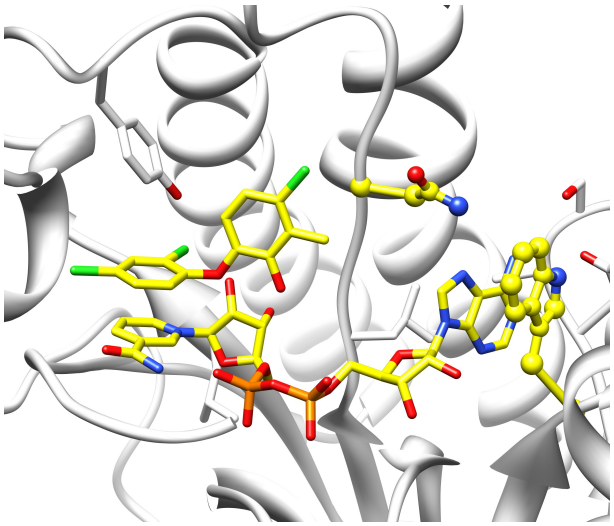


Figure 3.4a Crystal structure of PfENR with triclosan analogs of the 6-position substitution: compound **12**.

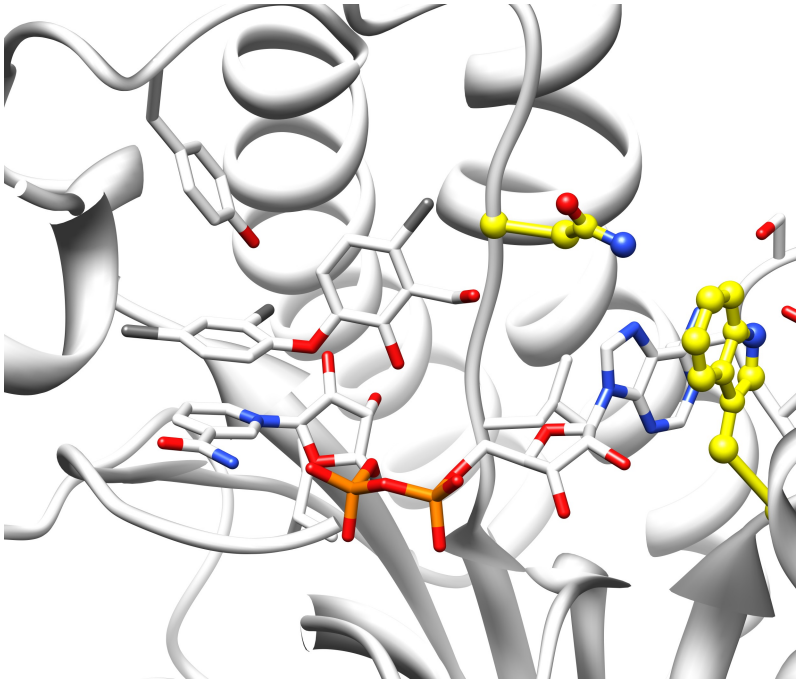


Figure 3.4b Compound **14**.

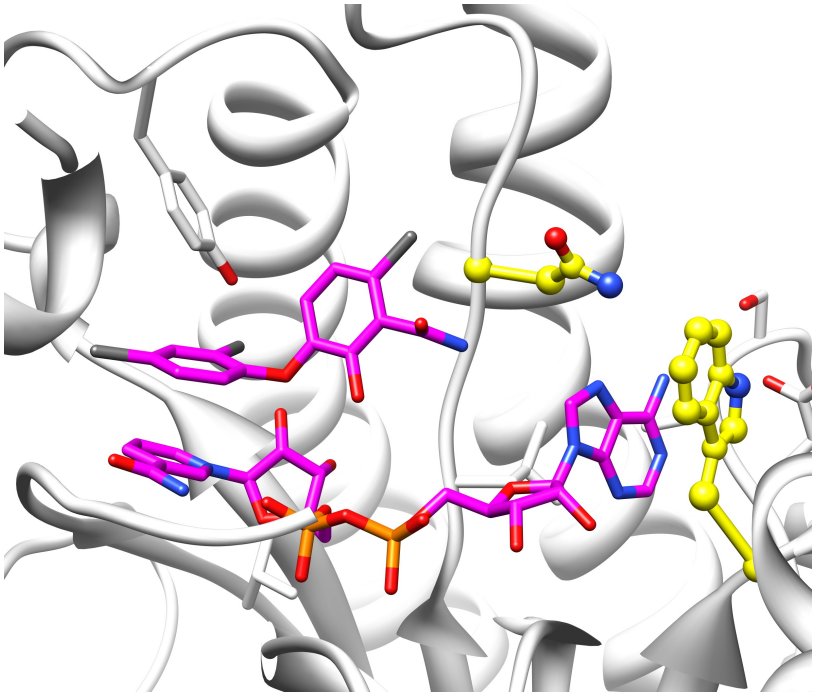


Figure 3.4c Compound **23**.

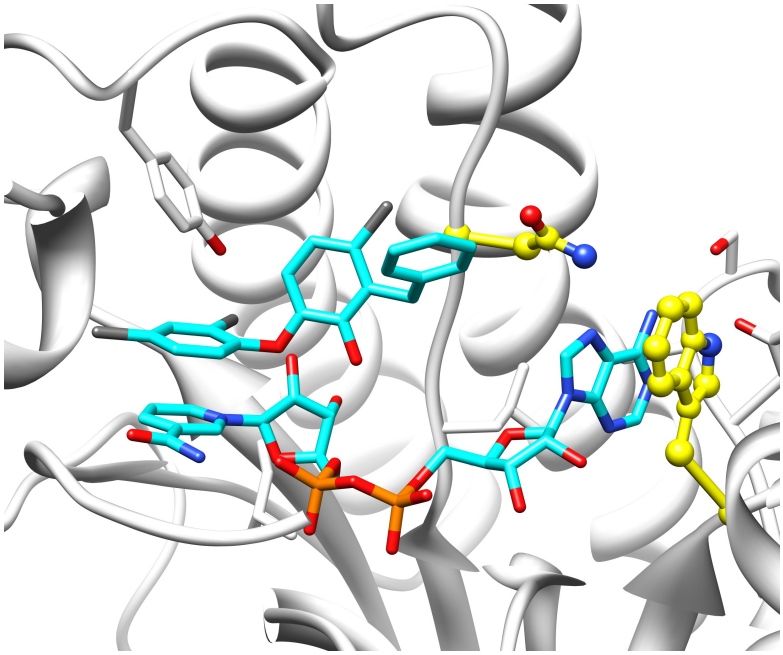


Figure 3.4d Compound **43**.

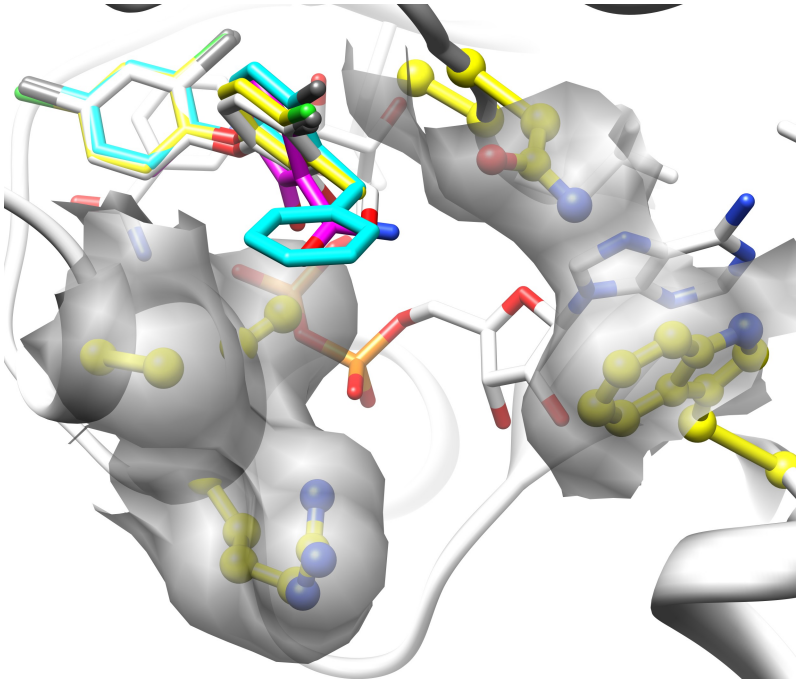


Figure 3.4e Overlap of 6-subsitution analogs binding in the flipped orientation.

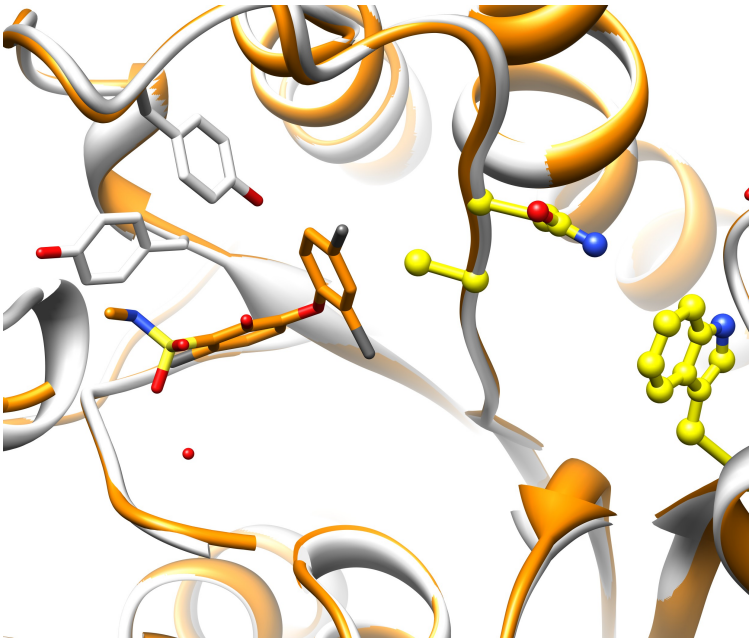


Figure 3.4f Compound **32**.

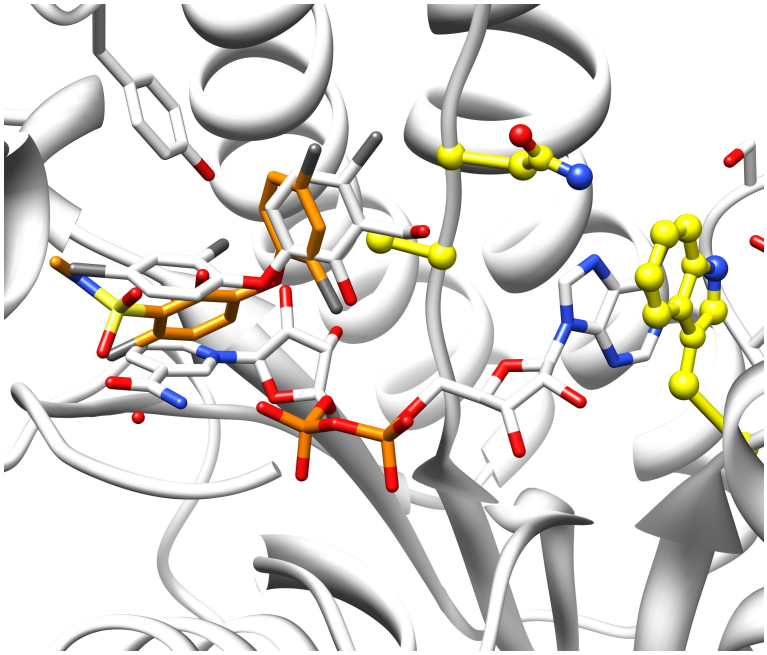


Figure 3.4g Ring A of compound **32** blocks NAD^+ binding.

The crystal structures of these derivatives are probes for revealing the interactions that could contribute to the inhibitor affinity at this NADH access tunnel. In compound **23** (the carboxamido group), the amine group interacts with the oxygen from the carbonyl group in the backbone of Ala217, and it forms a hydrogen bond at a distance of 3.1 Å. Compounds **12** and **43** use the methyl and benzyl group to form van der Waals interactions to the side chains of Ala319, Ala322, and Val222, and these two compounds are the most potent inhibitors among the 6-position substituted derivatives. Compound **14** (formyl group) is a potent inhibitor with a non-bulky modification, which is not expected to cause any penalty for binding. The crystal structures explain the two-rings-flipping phenomenon, because there is no available space at the original 6-position direction to accommodate additional modification. All the space at this direction is occupied by the side chain of the Tyr267, which prevents the binding of these 6-substituted analogs and results in a flipped binding conformation. Interestingly, one of the weak inhibitors, compound **32**, binds in the same orientation as the original one in the triclosan structure (Fig 3.4f). However, the electron density for the α7 (residues 318–325) is missing because this region becomes completely disordered on compound **32** binding. Due to the narrow space between inhibitor and Tyr267, compound 32 binds in a conformation with a flipped ring A. The 1-hydroxyl group on the ring A flips 141°, measured from the 1-hydroxyl of triclosan to the 1-hydroxyl of compound **32** with a center at the 1-carbon in the ring A, and the flipping ring disrupts the critical hydrogen bond between inhibitor and the hydroxyl of the Tyr277 side chain. In this binding conformation, the ring A is placed where the nicotinamide moiety of NADH and ring A

of triclosan bind (Fig 3.4g). Therefore, NADH is not seen in this structure. The oxygen of the sulfonyl group forms a hydrogen bond to a water molecule at a distance of 2.5 Å, and this water molecule also forms a hydrogen bond with the oxygen of the carbonyl group of Leu315 at a distance of 2.3 Å. The structure of compound **23** suggests that designing a substitution at the 6-position might cause either swapping the positions of ring A and ring B, or the ring A flipping. Both conformations result in losing the hydrogen bond to Tyr277, which explains why the inhibition by the 6-position derivatives is not as potent as triclosan.

The inhibitor binding site defined by the triclosan analogs with substitution at 2'-, 4'-, 5-, and 6- position

The inhibitor-binding site was surrounded by helix α7 (Pro314–Lys325) and α8 (Tyr366–Tyr375) at the front, and two loop regions (Ala217–Gln223 and Tyr267–Met281) at the back, surrounding where NAD binds. The portal of this inhibitor/substrate binding site to the solvent-accessible surface is on the side between loop (Ala217–Gln223) and α7 (Arg318–Ile323). The shape of the entrance is narrow with a size of 12 Å × 6 Å. Most inhibitors bind at the upper part of the portal, which is close to Val222 and Ala322. The lower part of the portal near Arg318 and Asn218 is where CoA binds. At the center of the binding site is Tyr277. It is evident that the formation of the hydrogen bond to Tyr277 is crucial for substrate or inhibitor binding. The inner part of the binding pocket is hydrophobic and is surrounded by Tyr267, Ile323, Ala320, Phe368, Pro314, and Ala372 right above the nicotinamide moiety of the

NAD, which is where the 5-chloro group from the triclosan is placed in the binding site. We found that the bulky substituted modification at this binding direction can lead to conformational changes to open up to two new hydrophobic interaction sites. The lower site is the one that displayed isoniazid-NAD adduct binding. This site is surrounded by Tyr267, Pro314, and Ala372. The upper new interaction site is formed on a flipped Phe368, which is caused by the larger moiety substituted at the 5-position of ring A on triclosan.

New chemical backbone templates discovered by compound library screening

Although triclosan is a potent inhibitor of PfENR, the majority of triclosan gets glucuronated in the human liver. The enzyme UDP-glucuronosyl transferase catalyzes this reaction by transferring a glucuronosyl moiety to a substrate that contains hydroxyl and amine groups. In this case, the 1-hydroxyl from triclosan becomes a glucuronidated phenolic hydroxyl moiety, which will result in a drastic decrease in the affinity to PfENR due to the loss of that critical hydrogen bond to the hydroxyl of Tyr277. Consequently, glucuronated triclosan is not a potent inhibitor of PfENR, and a new compound with a new chemical template is needed for inhibitor optimization. From the SAR study of the triclosan analogs, compounds with 5-substitution showed low IC_{50} in the nanomolar range and induced a conformational change at the side chain of Phe368. This side chain flipping opens up a new site for optimizing the inhibitor affinity. Therefore, compound libraries containing 55,000 structurally diverse compounds were

tested using a high-throughput screening method. Among the libraries, 880 compounds from the Prestwick Chemical Library were FDA-approved drugs. Another 2,000 compounds from the MicroSource spectrum collection were marketed drugs and natural products. The Tuberculosis Antimicrobial Acquisition and Coordinating Facility (TAACF) Diversity set contained 1,200 compounds. A commercial compound library from ChemDiv for TB testing also was screened. The results of the high-throughput screen were filtered by criteria including an IC_{50} below 1 uM, divergence from the FAS-I inhibitors, and compound solubility. A few chemical templates, which had the potential for the design of higher potency inhibitors and good correlation between *in vitro* and *in vivo* studies, were selected. The enzymatic screen of PfENR against their chemical libraries led to the identification of three series of platforms (piperidines, indoles, and tetrahydroisoquinolines) (Table 3.2). Each one of them was picked to target specific regions in the active site of PfENR, to pursue the SAR study within each group, and to validate the key residues for improving inhibitor-binding affinity. The first series is piperidine, which aimed to reach the buried Tyr267 to form a hydrogen bond. Then, the indole series was picked to reach the isoniazid-binding pocket by modifying the substitution around the indole ring. The tetrahydroisoquinolines were selected based on forming a π-π interaction with NAD^+. In all three series, the hydrogen bond to the hydroxyl of Tyr277 is a conserved interaction. This hydrogen bond also is crucial for the inhibitor affinity. To preserve this interaction between inhibitor and protein, an acryl amide was selected to be the central unit of all compounds with new backbone template.

Table 3.2 Chemical structure and IC_{50} of the compounds with the new backbone.

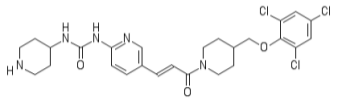
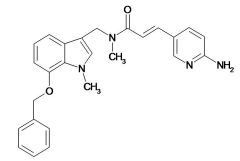
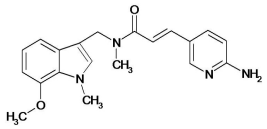
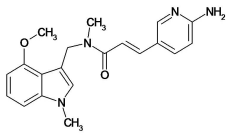
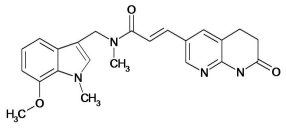
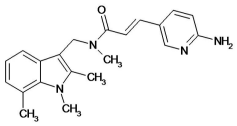
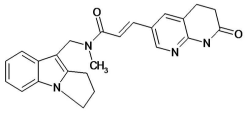
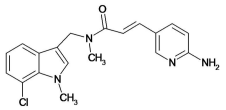
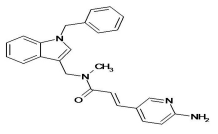
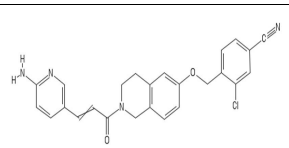
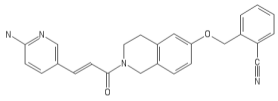
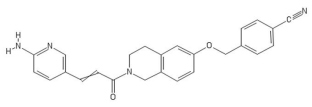
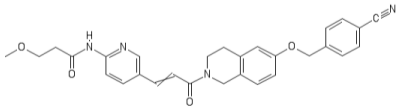
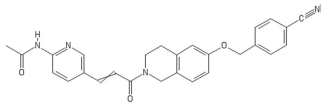
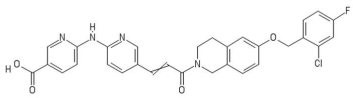
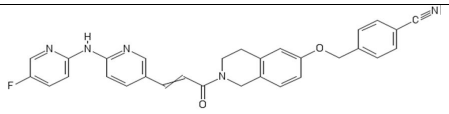
piperidine		**PfENR IC_{50} (μM)**	**Pf IC_{50} (μM)**
325		0.006	2.1
Indole		**PfENR IC_{50}**	**Pf IC_{50}**
313		0.54	17.6
059		1.0	n/a
160		0.61	n/a
363		0.92	n/a
813		0.41	n/a
865		0.41	n/a
914		NA	n/a
023		14	n/a

Table 3.2 Continued

THQ		PfENR IC_{50}	Pf IC_{50}
698		0.005	0.72
853		0.09	1.66
855		0.29	3.75
987		0.092	5.8
696		0.085	3.12
745		0.028	n/a
146		0.064	0.12

The complex structure of PfENR and a piperidine compound

The goal of this piperidine series is to make an interaction with Tyr267 that has been shown in the structure of the isoniazid:NAD adduct (PDB:2NQ8). One compound **325** from the piperidine series has high affinity to PfENR and the IC_{50} value is 5 nM. The crystal structure reveals a central moiety, an acryl amide fused in the piperidine ring, and resides at the region that form a hydrogen bond to the hydroxyl group of Tyr277 at a distance of 2.4 Å (Fig 3.5). The piperidine ring stacks with the nicotinamide moiety from NAD+. The other major hydrophobic interaction is at the portal of the active site, which is formed between Val222 and the pyridine ring of compound **325** at a distance of 4.3 Å measured from the gamma-carbon of Val222 to the 4-carbon on the pyridine ring. The urea moiety linked to the pyridine ring formed two hydrogen bonds to the backbone oxygen of Ala219 with a distance of 2.9 Å, and 3.4 Å to the two nitrogens on the urea moiety. On the other side of compound **325**, the tri-chloro-benzyl-methoxyl group is located at the hydrophobic site inside the active site, and has hydrophobic interactions with Tyr371, Phe368, and Ile323. The two chloro groups (2-chloro and 4-chloro) form chloro-arene interactions to the aromatic ring of Tyr267 and Tyr371. The distance from the 2-chloro to the aromatic ring of Tyr267 is 3.2 Å, and it is 3.4 Å from the aromatic ring of Tyr371. The interactions between the tri-chloro-benzyl group to protein residues of Tyr267, Tyr371, Val274, and Phe368 suggest that the hydrophobic modification to reach these three residues will increase binding affinity.

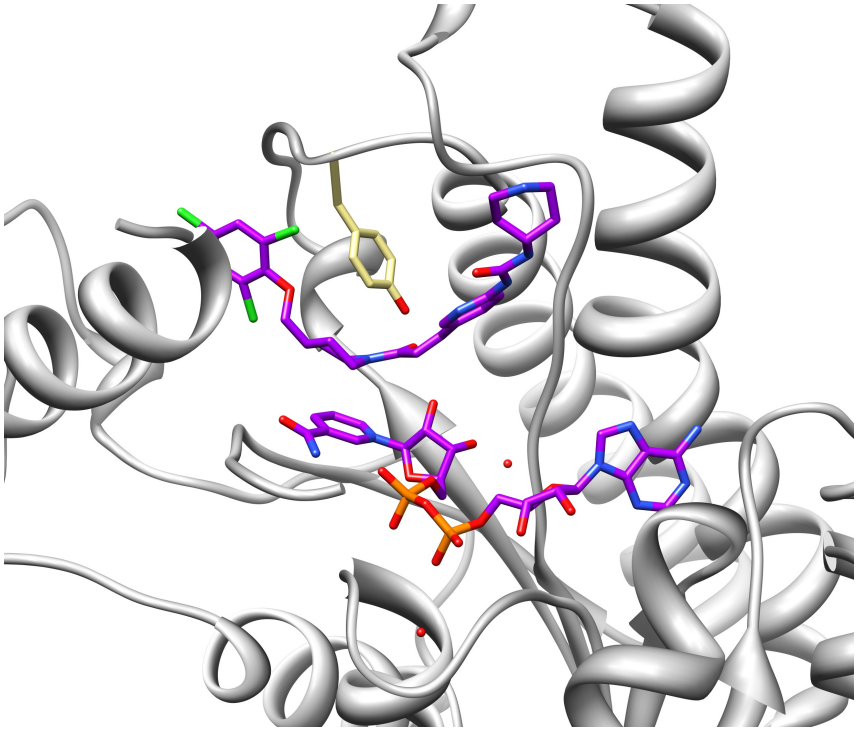


Figure 3.5 Binding of the piperidine compound **325** in the active site.

The complex structures of PfENR and the compounds with indole platform

For the eight compounds (**313**, **059**, **160**, **813**, **023**, **914**, **363**, and **865**) from this indole series, the acryl amide moiety, which is the moiety at the center of the molecules, is kept to form a hydrogen bond to the hydroxyl from Tyr277. The conserved indole group is attached to the nitrogen of the acryl amide moiety. Six compounds (**313**, **059**, **160**, **813**, **023**, and **914**) have a 2-aminopyridine group linked through the -ene group of the acryl amide moiety. The other two compounds (**363** and **865**) have a 3,4-dihydro-1,8-naphthyridin-2-one group attached at this position on the acryl amide moiety. The 2-aminopridine and naphthyridinone groups are designed to reach Asn218 and Ala219. Hence, the indole group will lie inside the active site and the modification at the position

of N-, 2-, 4-, and 8- could reach the hydrophobic pocket, constituted by Tyr267, Ile367, Phe368, and Ile323. The IC_{50} values for the indole series are in the range of 0.4 μM to 14 μM. Except for the compound **023** with an IC_{50} of 14 μM, the rest of the indole compounds are at 0.4 μM to 1 μM. Compounds **813** and **865** show the lowest IC_{50} of 0.4 μM (Table 3.2).

The crystals were obtained by co-crystallization. Each compound was incubated with protein and screened with a grid of 48 conditions around the hit condition. Data collection and structural determination were as optimized previously, and the crystals had the same packing with identical space group and unit cell as the crystals of PfENR:NADH. The indole compounds all bind in the same orientation and the indole ring, as proposed, lies in the hydrophobic pocket and forms a T-shaped aromatic ring interaction to Tyr277. Except for the indole ring of compounds **160** and **023**, all the indole rings are at the same orientation and the N-methyl moiety on the indole rings has hydrophobic interactions with Ala372 and Val274. The distance of the N-methyl to the γ-carbon on Val274 and β-carbon on Ala372 is 4.4 Å and 3.7 Å, respectively. The 7-position modifications (7-methoxyl, 7-methyl, and 7-chloro) on the indole ring, shown in the structure of compound **059** (Fig 3.6b), **363** (Fig 3.6f), **813** (Fig 3.6d), and **914** (Fig 3.6h), point at the direction of Phe368 and form hydrophobic interactions with the aromatic ring of Phe368. The distance is 3.8 Å, measured from the carbon on the 7-methoxy of compound **059** to the γ-carbon of Phe368. The 2-methyl modification in compound **813** reaches the isoniazid binding site that is approached by the isoniazid-NADH adduct. Here, the modification on the 2-position could form an interaction with

the aromatic ring of Tyr267. Due to the inhibitor binding conformation, the modification at the 4-position induces the indole ring to flip by 180°. This results in the N-methyl moiety of compound **160** pointing at Val222 and Ile323 (Fig 3.6c). The distance from the carbon of N-methyl to the γ-carbon of Val222 is 3.4 Å, and Ile323 interacts with the indole ring at a distance of 3.1 Å from the δ-carbon to the 1-nitrogen from the indole ring. The 4-methoxy moiety lies in the isoniazid-binding site. Compound **023** also shows a flipped indole ring (Fig 3.6e). Therefore, the N-benzyl moiety binds to where Ile323 is originally located and causes conformational changes at the α7 (residues 319–324), which is the reason why residues 322–325 become disordered and have no electron density. The other bulky modification is made in compound **313** with a benzyl-methoxy moiety at the 7-position on the indole ring. This bulky substitution does not induce indole ring flipping (Fig 3.6a). Instead, the 7-benzyl methoxy moiety reaches the hydrophobic site to interact with Phe368 and Tyr371.

The analysis of the correlation between the IC_{50} value and the inhibitor structures shows that bulky substitution at N-position and 4-position induces indole ring flipping. On binding, the bulky moiety on the indole ring (such as that of compound **023**) de-stabilizes the secondary structure of helix-7 and decrease the inhibitor potency. However, the bulky moiety modified at the 7-position (such as that in compound **313**) is able to reach the hydrophobic site of Phe368 and Tyr371, and keep the IC_{50} value similar to compounds **813** and **865**, which are the most potent inhibitors of this series.

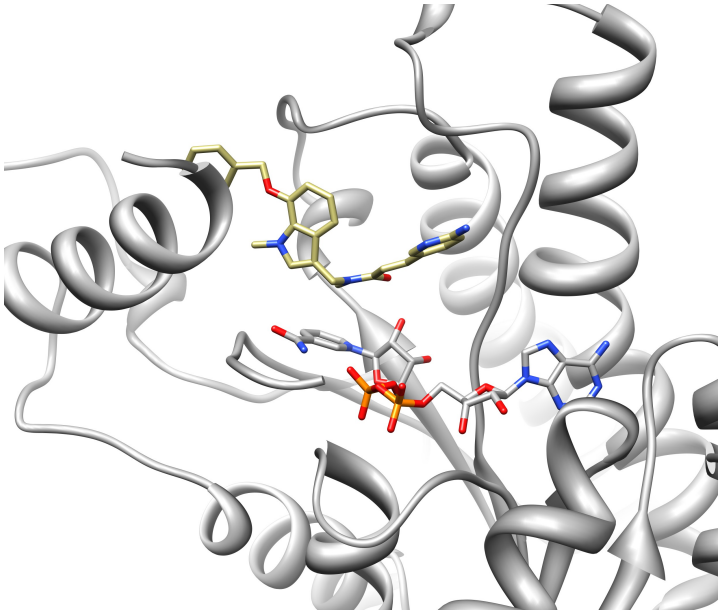


Figure 3.6a Crystal structure of indole compound binding in the active site: indole compound **313**.

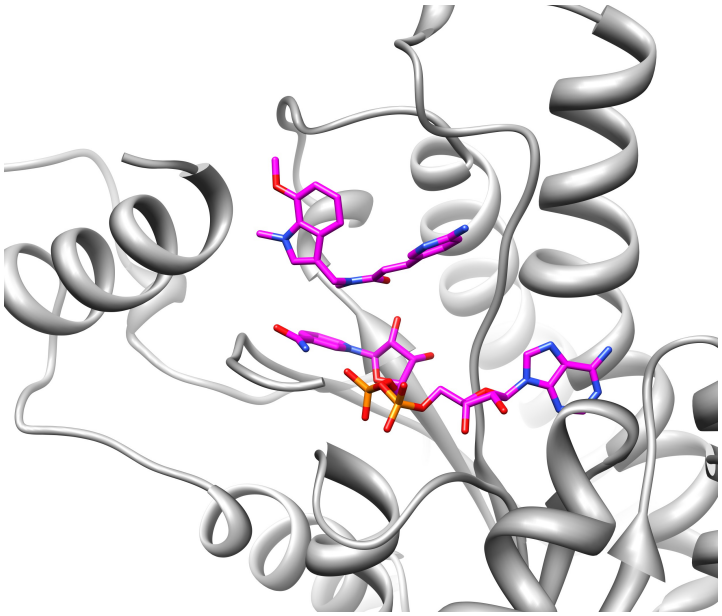


Figure 3.6b Indole compound **059**.

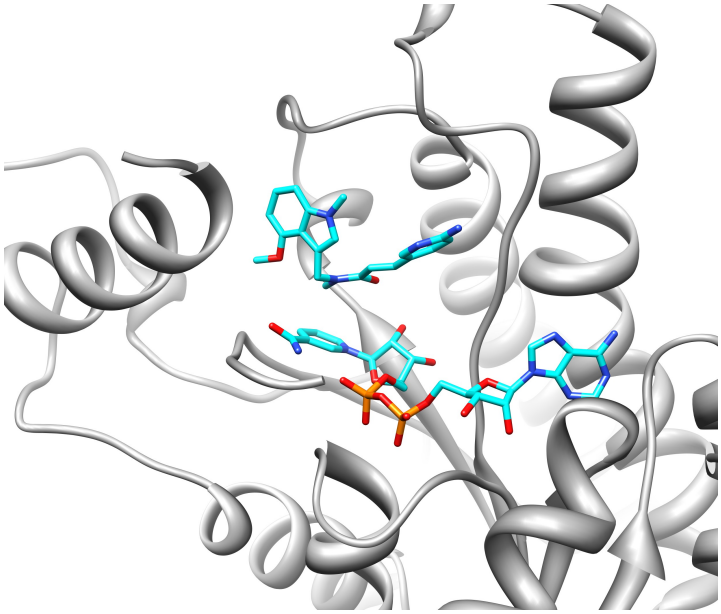


Figure 3.6c Indole compound **160**.

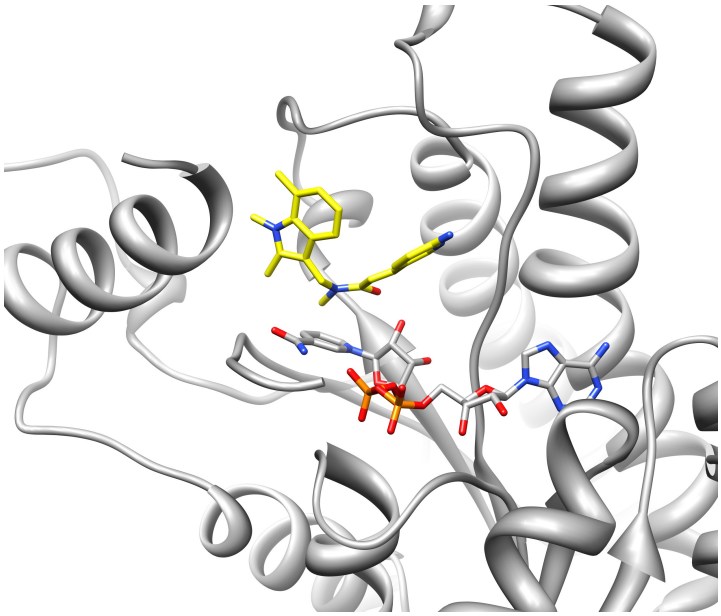


Figure 3.6d Indole compound **813**.

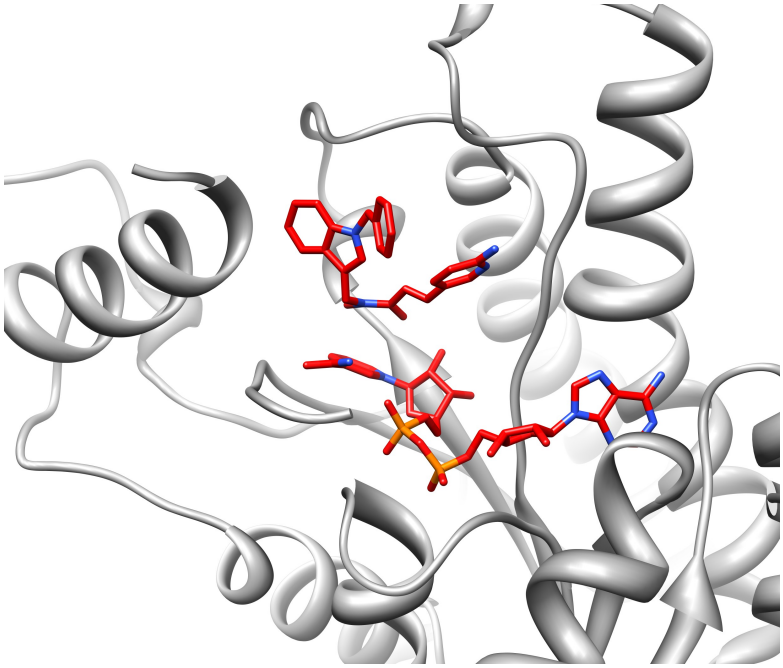


Figure 3.6e Indole compound **023**.

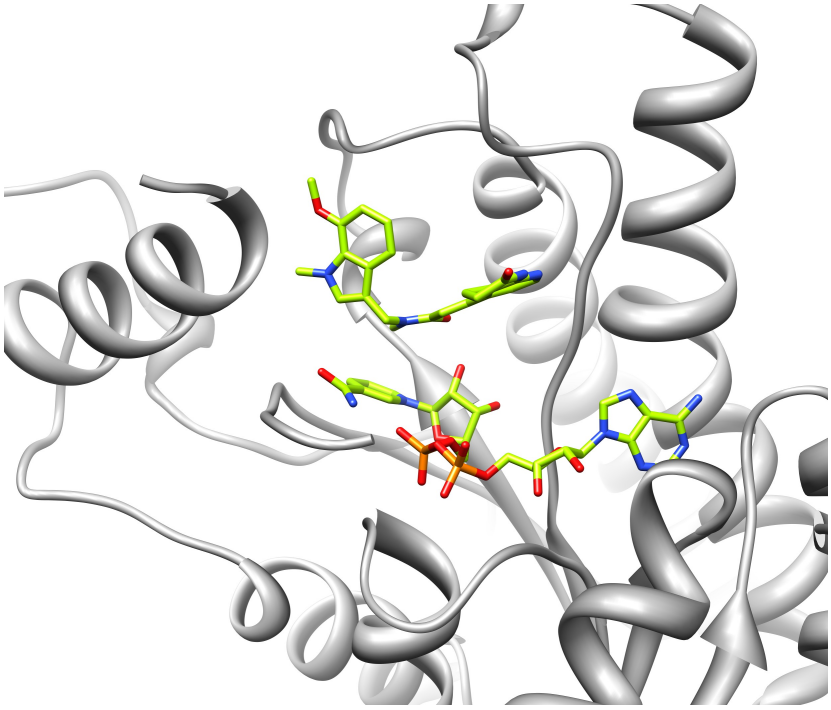


Figure 3.6f Indole compound **363**.

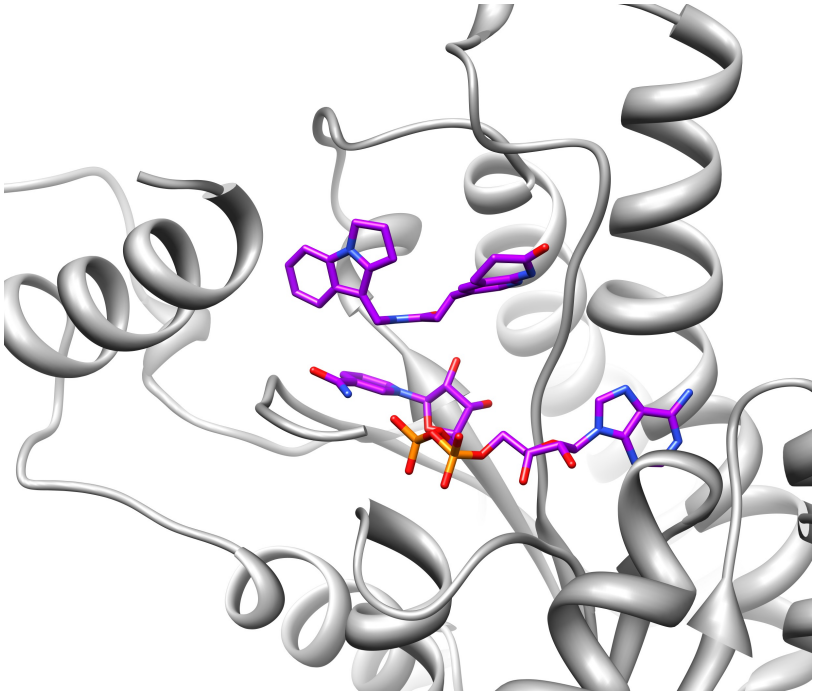


Figure 3.6g Indole compound **865**.

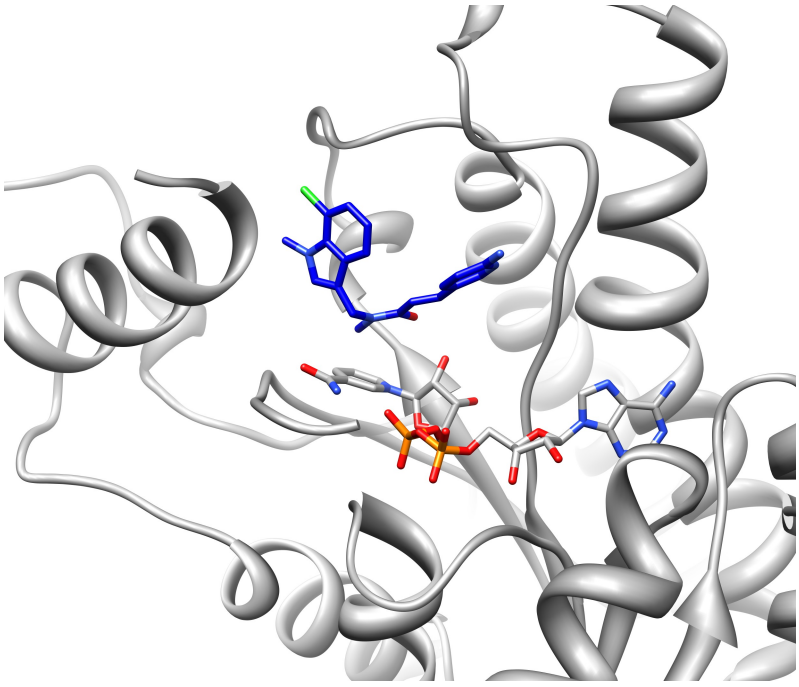


Figure 3.6h Indole compound **914**.

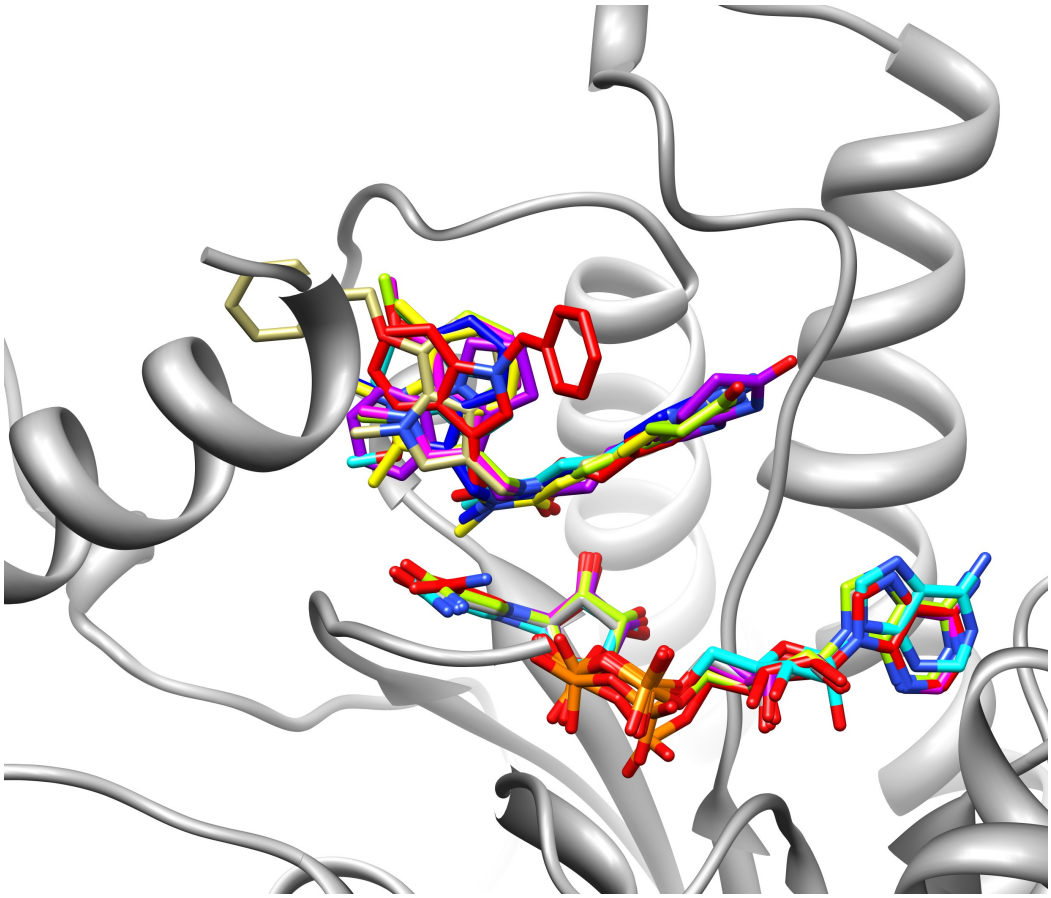


Figure 3.6i Indole compounds overlap.

The complex structures of PfENR and compounds with tetrahydroisoquinoline platform

The goal of the tetrahydroisoquinoline (THQ) series is to reach the hydrophobic site that has been explored with the indole series, and to further optimize the inhibitor potency. In this series, the center template base is still an acryl amide moiety, which is now fused to a tetrahydroisoquinoline moiety on the nitrogen of the acryl amide moiety (Table 3.2). To explore the hydrophobic site and the entrance portal, compounds with THQ moiety are elongated with another two rings. Therefore, the THQ compounds have at least three ring systems, including the center moiety of tetrahydroisoquinoline, pyridinyl, and benzyl rings. Compounds in this series all have a benzyl-methoxyl moiety connected at the 6-position of the tetrahydroisoquinoline moiety. The modifications on the benzyl ring, including cyano- substitution (**698**, **853**, **855**, **987**, **696**, and **146**) and halogen-substitution (**698** and **745**), are used to define the potential residues at the hydrophobic site that make interactions with inhibitors. On the other side of the compound, the 3-carbon on the acryl amide (2-propenamide) moiety is linked with a 2-aminopyridine moiety as a base for further modification. Four compounds (**987**, **696**, **745**, and **146**) are designed with further modification on the 2-aminopyridine ring such as the attachment of the acetyl (**696**), 3-methoxypropanamino (**987**), and pyridinyl (**745** and **146**). Amazingly, the compounds in this series have IC_{50} in the range of 5–288 nM. First, compared to all the triclosan analogs and the indole series, this tetrahydroisoquinoline series has the lowest IC_{50} on average. The IC_{50} of compounds **745** and **698** is 28 nM and 5 nM, respectively, which is lower than the IC_{50} of triclosan.

The crystal structures with THQ compounds were obtained through co-crystallization by the optimized-grid screening method. Crystals were produced in the same space group and were indexed to the same unit cell as those of PfENR:NAD. Therefore, the crystal structure were solved by using the PfENR structure with triclosan bound (PDB:1NHG) using a refinement program (rigid body program in REFMAC from CCP4 suite). Out of seven THQ compounds, six crystal structures of PfENR-THQ complex were obtained. The binding orientation shows that the 2-aminopyridine moiety locates at the entrance portal and the THQ ring forms a T-shaped aromatic interaction with Tyr277, which results the benzyl ring binding in the hydrophobic site (Fig 3.7). The hydrogen bond between the oxygen from the acryl amide and the hydroxyl from Tyr277 is still conserved in this series. Compounds **698** (Fig 3.7a), **855** (Fig 3.7c), and **853** (Fig 3.7b) have identical 2-aminopyridine ring binding at the entrance portal. The nitrogen from the 2-amino moiety forms a hydrogen bond to the oxygen of Asn218 at a distance of 3.4 Å, and the nitrogen from the pyridine ring forms a hydrogen bond to the nitrogen on the backbone of Ala219 at a distance of 2.9 Å. The pyridine ring has van der Waals interactions with Ala222, Ala322, and Ile323 at a distance range of 3.6–4.2 Å. One of the modifications on this pyridine ring, 5-carboxyl-pyridinyl moiety (**745**) seems to form extra interactions with Ala322 and Gln223, where it forms hydrogen bonds with the oxygen on the carboxyl moiety (Fig 3.7f). One is at a distance of 3.2 Å from the oxygen on the backbone of Ala322, and the other is at a distance of 3.0 Å between the nitrogen on the side chain of Gln223 to the oxygen of carboxyl moiety.

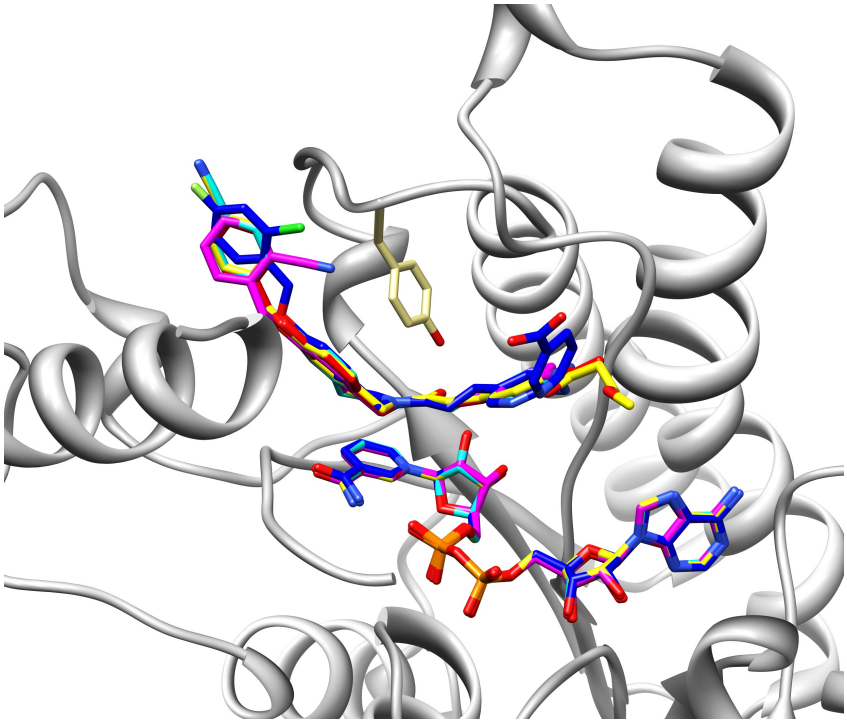


Figure 3.7 Crystal structures of the THQ compounds binding in the active site.

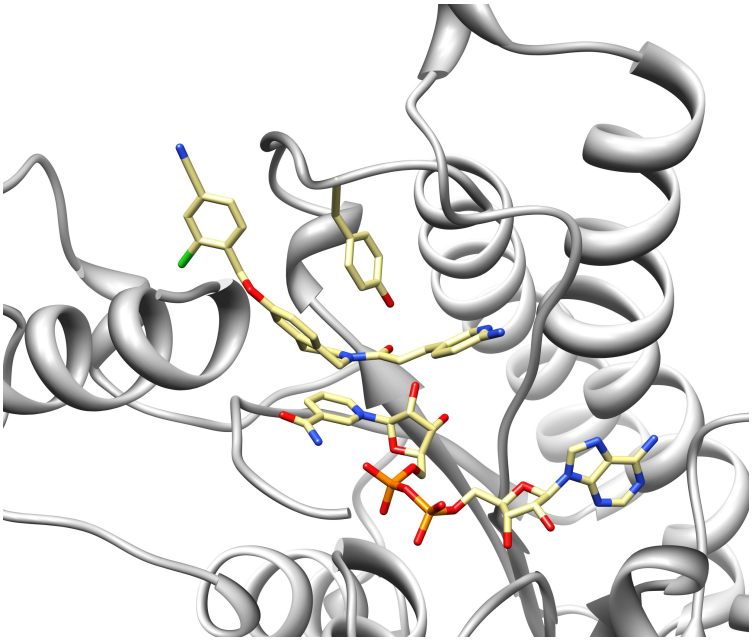


Figure 3.7a THQ compound **698**.

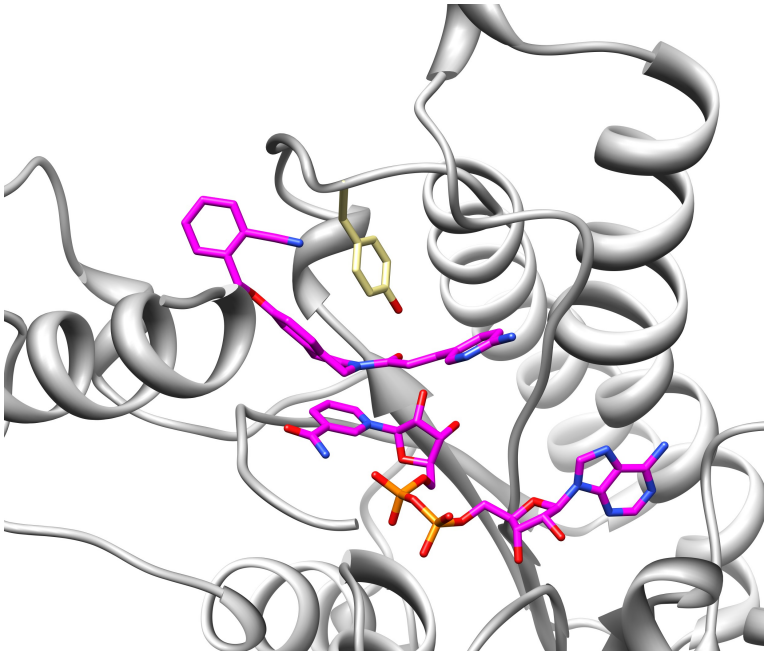


Figure 3.7b THQ compound **853**.

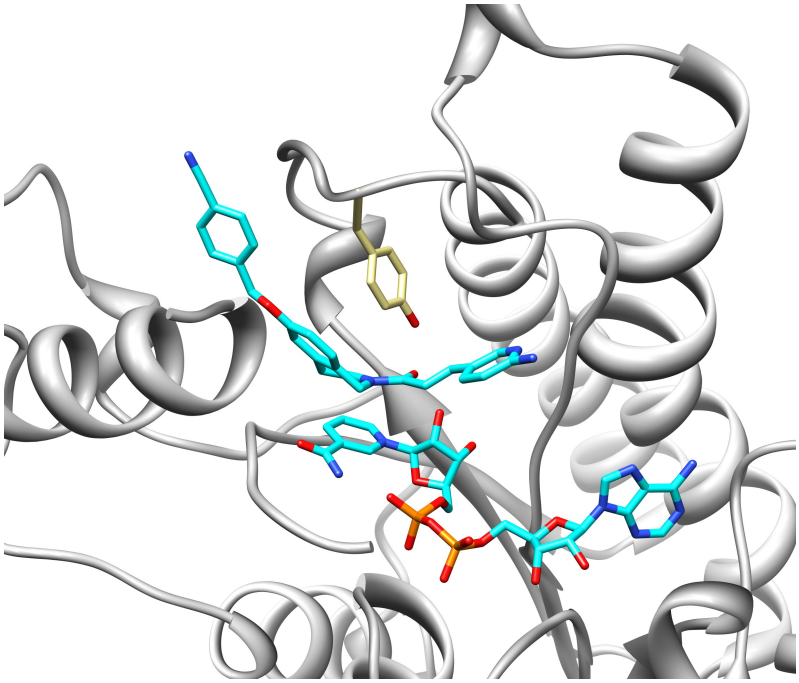


Figure 3.7c THQ compound **855**.

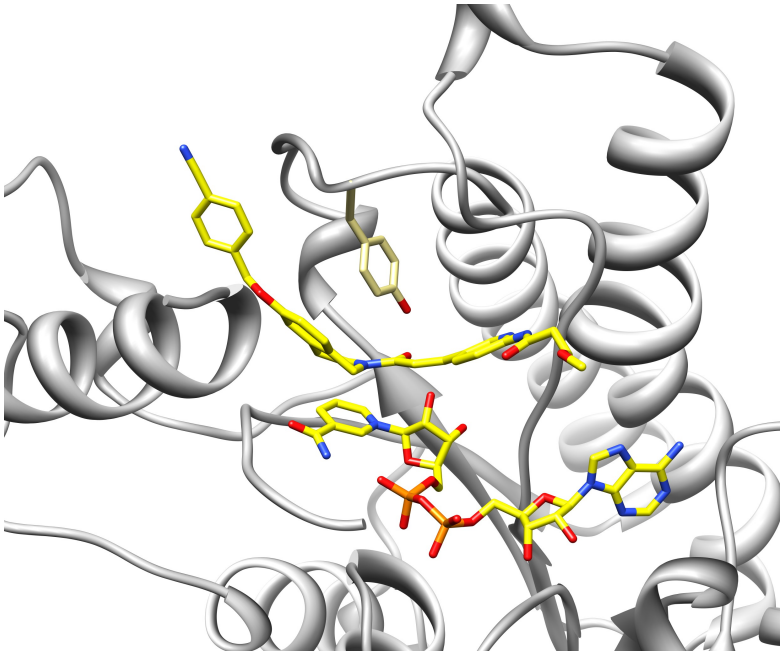


Figure 3.7d THQ compound **987**.

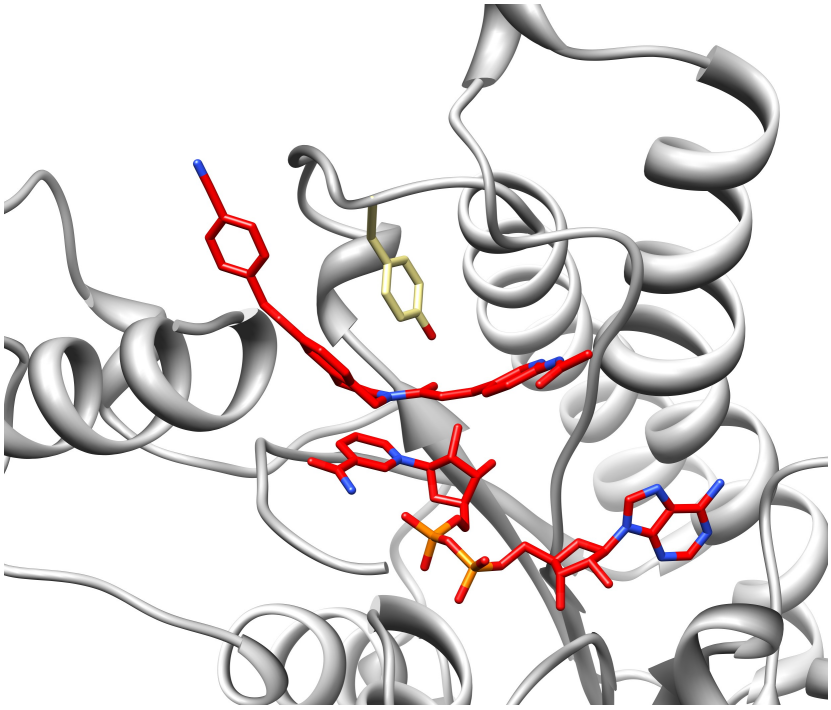


Figure 3.7e THQ compound **696**.

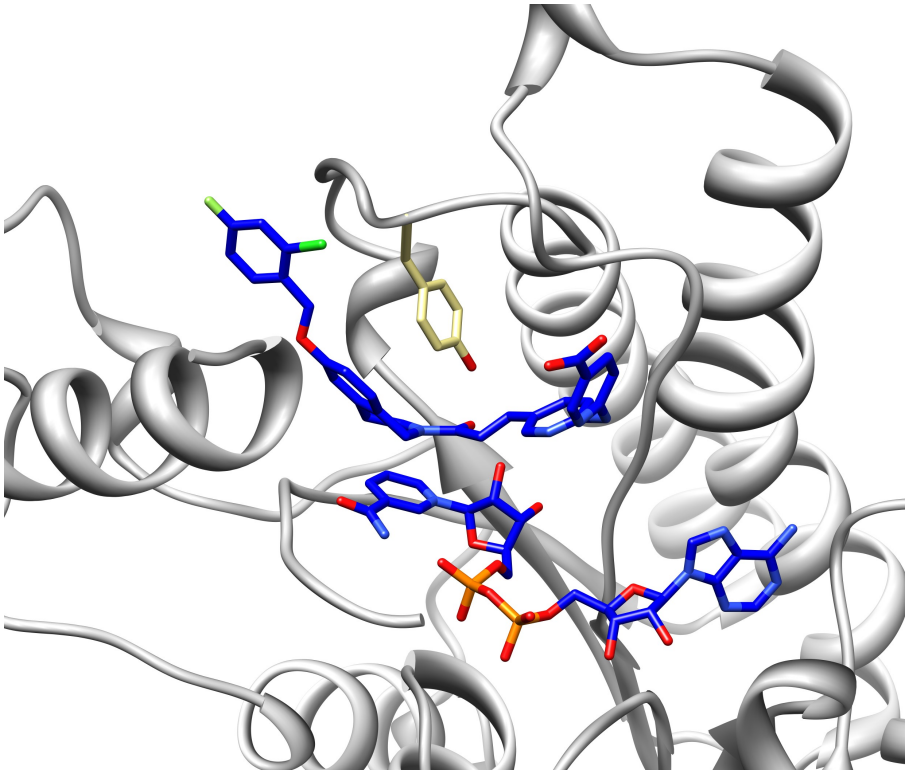


Figure 3.7f THQ compound **745**.

The side chains of Phe368, Tyr371, Pro314, Val274, and Ile426 from the adjacent protein molecule make van der Waals interactions to the benzyl ring. The interactions from the adjacent molecule also are observed in the structure of compound **313** from the indole series. Compounds **855** (Fig 3.7c), **987** (Fig 3.7d), and **696** (Fig 3.7e) all contain a cyano moiety attached to the para position of the benzyl ring. This cyano moiety points to Asp424 from the adjacent molecule, and the cyano moiety likely interacts with the carbonyl on the backbone of Asp424 at a distance of 3.4 Å, measured from the nitrogen of the cyano to the oxygen from the carbonyl of Asp424. Residues from Asp425–Glu432, as a part of the C-terminal loop, were missing in the previously

determined crystal data. When PfENR co-crystallized with **987** and **696**, these residues form a new helix that had not been seen before, which interacts with the cyano-substitution on the benzyl moiety (compound **987** and **696**). The modifications on the ortho position of this benzyl moiety in compounds **745**, **698**, and **853** show interactions with residues of Phe368 and Tyr267. The chloro substitution from compound **698** interacts with the hydroxyl of Tyr267 at a distance of 3.4 Å from the chloro to the oxygen of the Tyr267 hydroxyl. Yet, the chloro substitution of compound **745** forms a chloro-π interaction with Phe368. A cyano substitution on the ortho position from compound **853** binds at the same position as the chloro substitution from compound **745**, and has a CN-π interaction with Phe368.

The analysis of the correlation between the IC_{50} value and the THQ structures shows that the ortho substitution on the benzyl ring is crucial to make a compound such as **698** with an IC_{50} value of 5 nM. When comparing compounds **698** and **855**, the compound **855** without *o*-chloro substitution as the only structural difference from **698**, and its IC_{50} is 288 nM. Even with an *o*-cyano substitution (**853**), the affinity for PfENR is not as strong as that of **698**. This highlights the interaction between the chloro to the hydroxyl group of Tyr267. Alternatively, the *o*-chloro is able to form the Cl-π interaction to Phe368, which increases the binding affinity of **745** with an IC_{50} of 28 nM. The *o*-substitution on the benzyl ring forms interactions with aromatic rings such as Tyr267, Phe368, and Tyr371, which also is supported by the structure of compound **325** and its low IC_{50} of 5 nM. However, modifications on the 2-aminopyridine ring are not as critical as those on the benzyl ring located in the hydrophobic site, because compound **696** and

745 with different modifications show similar values of IC_{50}. Despite the low contribution to affinity from the modification on this pyridine ring, one advantage for drug optimization is found by comparing **146** and **855**. Compound **146** with a bulky substitution (5-floro-pyridinyl group) has lower IC_{50} than **855**. This suggests that further modification on the 2-aminopyridine ring will not dampen the inhibition, but it could be used for improving compound solubility, which could lead to better IC_{50} *in vivo*.

3.3 Methods

Expression and purification of PfENR

The PfENR-pET28a vector was made previously in the lab and was transformed in Rosetta-2 BL21 (DE3) cells (Novagen). The cells were grown in Terrific broth at 37°C, and when the cell culture reached OD_{600} of 0.8, isopropyl-1-thio-D-galactopyranoside (IPTG) was added to a final concentration of 1mM for induction. Then, the culture was further incubated for 5 hours at 37°C to overexpress the protein. Cell pellets were spun down at 4,000 rpm for 30 minutes and resuspended in buffer A (20 mM Tris pH 8.0, 500 mM NaCl, and 50 mM imidazole). The suspended cells were disrupted using a French press, followed by centrifuging at 15,000 rpm for 1 hour to remove cell debris. The clear supernatant was filtered by 0.22 μ m syringe filter and passed through a nickel column (Pharmacia).

Inhibition study

The PfENR inhibition enzyme assays were done in a 96-well format by monitoring oxidation of NADH at 340 nm and reading by plate reader (POLARstar Optima). The pre-mix containing 50 nM PfENR, 400 μ M NADH, and 40 μ M NAD+ in buffer of 20 mM Tris pH 7.5 and 150mM NaCl was transferred in aliquots to a COSTAR 96-well plate. Inhibitors were added to the well, and the incubation was at room temperature for 30 minutes. The substrate, 300 μM crotonoyl-CoA, was added by injector from the plate reader (POLARstar Optima) to initiate the reaction. The IC_{50} was calculated from a dose-response plot of enzyme fractional activity as a function of inhibitor concentration.

Crystallization, data collection and structure refinement

Pure protein was used for setting up the hanging-drop and vapor-diffusion method, and was crystallized with NADH and inhibitors. One method to obtain protein-inhibitor complex structure is the soaking method, which produces crystals of the binary complex containing PfENR and NADH. The final concentration of protein is 20 mg/mL in buffer of 20 mM Tris pH 7.5 and 150 mM NaCl. Prior to setting up the crystallization plate, 4 mM NADH was added to the protein, and the well solution was made with 2.5 M $(NH_4)_2SO_4$ and 0.1 M sodium acetate pH 5.6. The cover slides were used to mix the drops with 2 μL of protein incubated with NADH and 2 μL of well solution. The cover slides on each well were sealed, and the drops on the slides were equilibrated against the well solution at 18°C with the crystallization plate placed in a vibration-free

environment. Once crystals were formed, 1μL of inhibitor dissolved in acetonitrile (1–5 mM) was directly added to the drops containing crystals for the soaking process. This might take 1–3 weeks to allow inhibitors to have enough occupancy in the active site. The other method is co-crystallization, in which 20 mg/mL protein was mix with 0.5–1 mM inhibitor on ice for one hour before setting up the crystallization plate. The plates had 48 conditions on a grid, with a gradient change in concentration of $(NH_4)2SO_4$ along the columns (from 1.9–3.0 M) and a gradient change of sodium acetate along the rows (from 0.085–0.115 M). The plates were set by a robotic system (Art Robbins) with the sitting-drop vapor-diffusion method, and equilibrated at 18°C. Crystals from these two methods were mounted in the cryoprotectant paratone for data collection. During data collection, the crystals were collected under cryostream at 120 K, and were rotated in step from 0° to 120–180° with an oscillation angle of 1°. The crystals from soaking and co-crystallization were isomorphous to those of the binary complex, and were crystallized in the space group of P43212 with unit cells of $a = b = 134.0$ Å, $c = 84.0$ Å. Each asymmetrical unit contains a dimer of PfENR.

Due to the protein-inhibitor crystals being isomorphous, the initial phase was obtained from the previously solved PfENR-NAD^+ structure by running rigid-body refinement in the REFMAC program (CCP4). Each data set was refined by CCP4 or PHENIX and viewed by COOT. The protein coordinates of the PfENR-NADH complex (PDB:1VRW) were used to calculate the $F_o–F_c$ and $2F_o–F_c$ electron density maps. Viewing with COOT, data with an average resolution of 2.5–3.0 Å had inhibitor fitted based on the $F_o–F_c$ map, and were subsequently refined with PHENIX.

4. FUNCTIONAL IDENTIFICATION AND STRUCTURAL STUDY OF TAIL FIBER PROTEINS FROM MYCOBACTERIOPHAGE BXZ1

4.1 Introduction

Bacteriophages are known as phage and belong to a group of viruses that can only infect bacteria[115]. Viruses are tiny particles (in the nanometer range) that carry genetic material in their core, and a protective protein coat surrounds the viral genome[116, 117]. These particles are the most abundant biological entities in the biosphere, which self-replicate through a wide spectrum of foreign organisms (animals, plants, insects, and bacteria)[118, 119]. They live in diverse environmental conditions and exist in a variety of forms that are defined by the host's preference, morphology, genome type, and lifestyle[118]. A virus that can infect bacteria is called a bacteriophage[119]. Based on the phage morphology and nucleic acid properties, 95% of bacteriophages are classified in the Caudovirales order[120, 121]. In this order, they are characterized by a double-stranded DNA (dsDNA) genome and a tail organelle. The phages in Caudovirales share similar structural organization with a polyhedral head (capsid) that contains linear dsDNA[117, 121]. The capsid is attached to a tail, and the apparatus on the end of the tail is constructed into a special system that is responsible for recognizing the host and penetrating the bacterial cell wall[117].

Phages have a host range to infect certain bacteria specifically. This infection is efficient and has two types: lysogenic and lytic infection. The lytic infection can be

defined by plaque formation[119, 122]. A single phage injects the genome into bacteria in order to engage the cellular mechanism inside the host to become suitable for the phage to propagate. This process of lytic infection is then followed by bacterial rupture, which allows the progeny phages to release[118, 123]. Among the bacteriophages, tailed phages are found to be predominant. All tailed phages from the order Caudovirales are subdivided into three families based on the tail morphology: *Myoviridae* (long contractile tail), *Siphoviridae* (long noncontractile tail), and *Podoviridae* (short noncontractile tail)[121]. The construction of a phage tail requires multiple proteins from a tail assembly pathway. This pathway is thought to add to the complexity and metabolic cost of building the tail structures. However, the tailed phages have a long history of co-evolution with bacteria. The prevalence of tailed phages implies that the tails must provide phages with a tremendous evolutionary advantage[117].

The functions of the phage tail are designated for host cell recognition, attachment to the targeted bacteria, and penetration of the cell envelope[124]. Hence, this extra organelle in bacteriophage, which has not been observed in any other viruses, is an important mediator that facilitates the ability to fulfill the infection at an early stage. The phage tail is built into a cylindrical shape and is constructed by proteins polymerized in a helical conformation. At the distal part of the tail, it encompasses the fixation structures and the penetration device that contains the baseplate, spike, and tail fiber. The tail fiber is constituted at the most distal part of the tail, which extends from the base plate[125, 126]. The function of the tail fiber is in host recognition by making the first contact with the specific bacteria[73]. Despite the conserved tail organelle

within all the bacteriophages in Caudovirales, the tail morphology and the structure of the tail fiber reveal a major discrepancy among the three families of the tailed bacteriophages. The difference in the tail morphology lies in the tail length, and it has been used for phage classification. Electron microscopic images of the phages reveal that the tail length of *Siphoviridae* and *Myoviridae* is much longer than that in *Podoviridae*, and the length might be related to host selection and infection capacity of a phage[127, 128]. The tail fibers from *Siphoviridae* and *Podoviridae* are constructed by a shorter peptide chain, and there is only one type of the tail fiber reported for each of these two phage families. By contrast, *Myoviridae* has two types of tail fibers. One type of tail fiber is the same as that observed in *Siphoviridae* and Podoviridae, and it is known as the minor tail fiber[129]. The other is known as the major tail fiber; it contains a unique structure, which features six major tail fibers that extend from the base plate[125]. The structure of the major tail fiber in *Myoviridae* displays a longer length than its own minor tail fiber and the tail fiber of the other phage families. The major tail fiber has an assembly of a long fiber with a kink in the middle, and the end of this major tail fiber is constructed of major tail fiber proteins (major TFPs)[130]. Clearly, this suggests the importance of the major TFPs in the early process of phage infection.

The major TFP located at the tip of the tail fiber makes the very first contact with a receptor on the surface of the bacteria. This protein plays an important role in providing specific recognition and tight binding to allow phage docking on the bacteria. The tail fiber initiates crucial binding to the bacteria in the process of the phage infection, and it is likely that this binding results in host specificity[124]. *Myoviridae*

utilize this major TFP to recognize bacteria while floating and randomly encountering organisms in the environment. Once the binding of the major TFP to the bacteria is established, it allows the phage to settle on the surface, and the kinks in the major tail fiber allow the minor TFP to make secondary contact onto the bacterial surface. Consequently, the tail spikes hold on to the surface and induce the conformational changes in the base plate[125]. The tail is contracted in order to penetrate the membrane and inject the phage genome into the bacteria[126]. Because the binding between the major TFPs and the bacteria is the required first step of phage infection, elucidation of the molecular structures of the major TFPs has been crucial for understanding bacterial recognition by bacteriophages. Any knowledge gained about the mechanism of recognition by the tail fiber proteins could be applied toward diagnosis and therapy of the diseases caused by bacteria.

Mycobacteriophages are known as phages that cause infections specifically in mycobacteria. It has been discovered that two different mycobacteriophages (I3 and bxz1) could be used as a tool for gene transduction in several species of mycobacteria[127, 131]. Phage I3 and bxz1 are lysogenic phages, which integrate phage nucleic acids into the bacterial genome; phage proliferation depends on bacterial reproduction [122]. They have shown to be capable of transducing auxotrophic and drug-resistance markers in *Mycobacterium smegmatis* (*M. smeg*). These transducing phages in mycobacteriophage are both found in *Myoviridae*, which contain the major tail fibers in the tail structure. Bxz1, which was first isolated from soil samples, has shown higher transduction efficiency than I3[131]. Particularly in mycobacteriophages, the tight

binding of the major TFP to the bacterial surface is likely to be one of the factors that result in a higher frequency of genetic transduction in mycobacteria. In our study, we have identified the gene encoding TFP in bxz1 by comparing the genetic sequences of TFPs from other bacteriophages such as T4 and TM4. Based on our hypothesis of the identified TFP gene, the binding of an engineered major TFP to the bacterial surface will demonstrate the bacterial targeting mechanism of the mycobacteriophage. Additionally, the crystal structure of the TFP from bxz1, reported here, is the first TFP structure of mycobacteriophages. The structure reveals a fold, which is observed in viral fiber protein, and exhibits a novel structure of the major TFPs from all bacteriophages.

4.2 Results and discussion

Identification of the tail fiber protein of mycobacteriophage

Due to the ability of efficient gene transduction and a specific host range to *M. smeg*, mycobacteriophage bxz1 was selected for identification of tail fiber protein (TFP). Mycobacteriophage Bxz1 is in the *Myoviridae* family, where only 1% of the mycobacteria are classified.[128, 132] This phage has a specific host range and exhibits specificity to *M. smeg* and *M. vaccae*, but not to *Mtb[131]*. The specificity between the bxz1 and *M. smeg* was used to guide the elucidation of the mechanism of phage recognition. Furthermore, *M. smeg* is a common model used in the lab for *Mtb*-related studies because of its short doubling time. The advantage of using the *M. smeg*-bxz1 model is that *M. smeg* is not pathogenic to humans, but is able to cause infection in mice. Therefore, the application can be expanded to test the therapeutic and diagnostic design

in a mouse model.

Although the genomic sequence of bxz1 has been assembled, the assignment of gene function has not yet been completed. To identify the gene that functions as TFP, the gene cluster of the tail assembly and the tape-measure protein were first explored. The *TFP* gene is commonly found with tape-measure protein in the operon of the tail that is observed in the gene organization of mycobacteriophage TM4 (Fig 4.1a) and lactococcus phage TP901-1. This tape-measure protein is a basic component to the tail assembly, and it is also a tool used to determine the length of the phage tail. For example, *gp17* is annotated as a tape-measure protein in *Siphoviridae* TM4. In the proximate region, the gene *gp14* that is found to be a putative TFP in TM4 is located within this tail assembly operon. In the genome of the lactococcus phage TP-901, the gene *orf45* is a tape-measure protein, which is in the tail assembly at the adjacent region to its tail fiber protein orf49. When the genome of bxz1 was newly sequenced, Pedulla *et al.* reported that *gp95* in bxz1 is a homolog to *gp17* in TM4. Proximal to *gp95*, the gene *gp112* is reported to be a TFP. This shows that bxz1 also follows the pattern of tail assembly gene clusters containing the genes encoded as tape-measure protein and TFP (Fig 4.1b). Based on the nucleotide length of *gp112* and the sequence alignment to that in other *Myoviridae*, this gene should be identified as the minor TFP in bxz1. However, the major TFP could not be found in the proximity of this tail assembly.

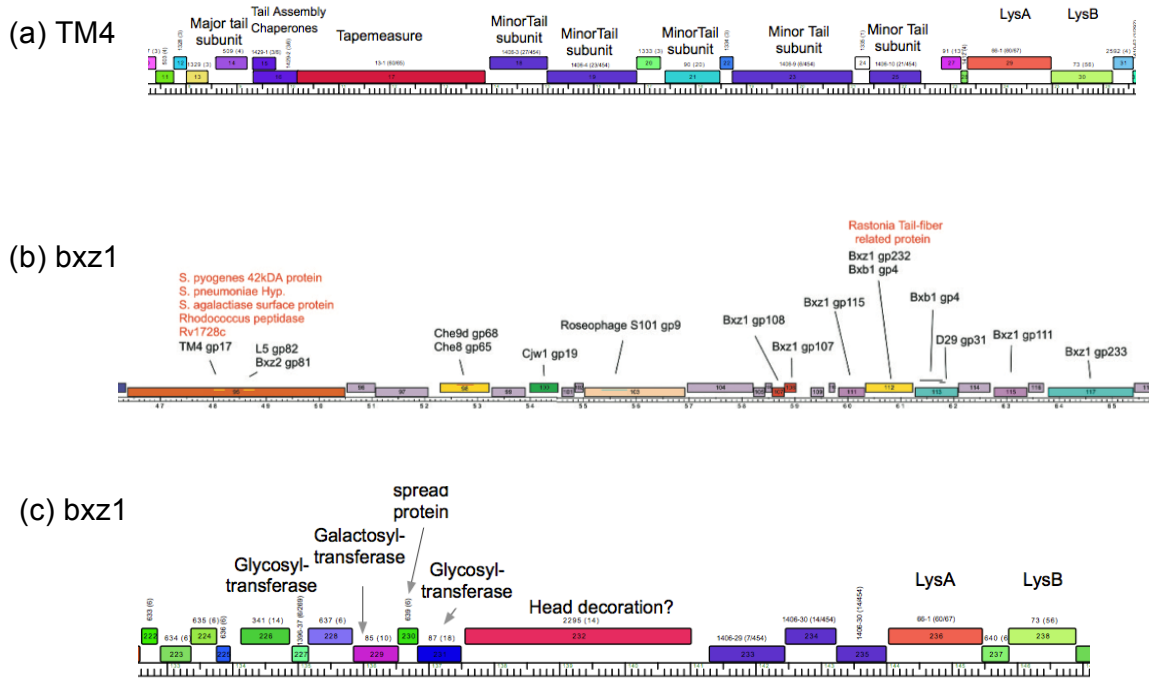


Figure 4.1 Gene organization of tail assembly in (a) TM4 and (b) bxz1. (c) The proximal region of the putative major tail fiber protein (TFP) in bxz1.[128]

There are no reports or clear indications of the major TFP in bxz1. The only information that could lead to finding the major TFP in the genome of bxz1 is that *gp112* was reported to be a homolog to the *gp232* within its own genome, even though *gp232* is not located in the tail assembly (Fig 4.1c). When looking into the gene organization of the major TFP (*gp37*) from enterobacteriophage T4, the TFP is not located in the tail assembly region where the tape-measure protein is typically found. Therefore, T4 and

bxz1 could have similarities in the gene organization of the major TFP, because both of them are in the same subclass of *Myoviridae*. The sequence alignment shows a sequence identity of 34% in the range of the residues 228–298 from *gp112* (301 residues in total) and residues 1081–1149 from *gp232* (1151 residues in total) (Fig 4.2a). These regions displaying aligned sequence are where the tips of the major and minor TFP individually form. The identified major TFP (*gp37*) and minor TFP (*gp112*) in T4 also have a similar pattern in their sequence alignment. Like the sequence similarity in gp112 and gp232, the residues 751–870 of *gp37* (1026 residues in total) and the residues 273–405 of *gp12* (527 residues in total) have a sequence identity of 28%, and yet these regions showing sequence similarity are not directly at the C-terminus (Fig 4.2b). Therefore, the similarity in the protein sequence at the C-terminus, which has been observed for translations of *gp112* and *gp232*, might not be conserved in all species from *Myoviridae*. However, it suggests that these major and minor TFPs of bxz1 might interact with similar receptors from *M. smeg*.

Figure 4.2a

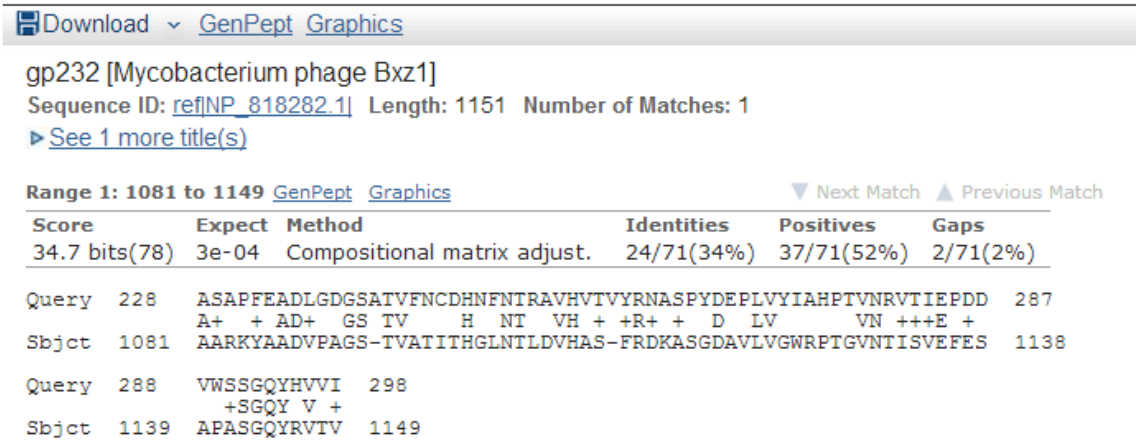


Figure 4.2b

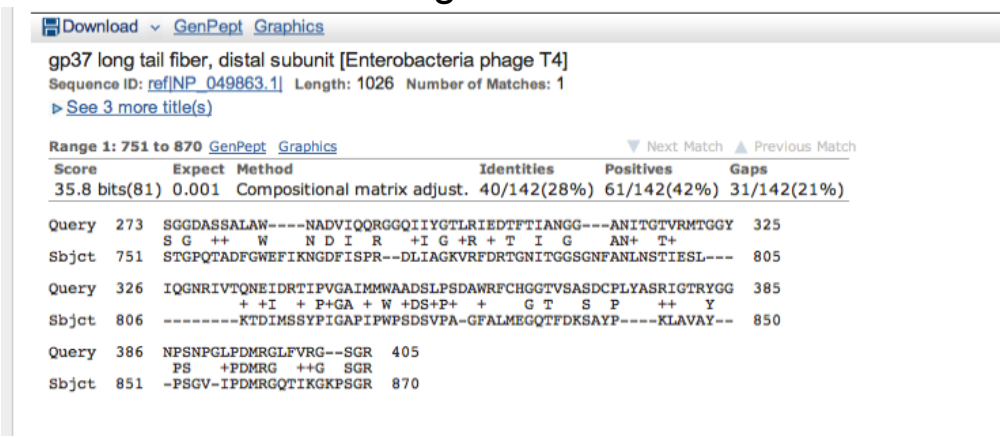


Figure 4.2 Sequence alignment of the major TFP and the minor TFP from two strains of *Myoviridae*. (a) gp232 and gp112 from mycobacteriophage bxz1; (b) gp37 and gp12 from enterobacteriophage T4.

The protein sequence of TFP, especially the C-terminal residues, provides the specificity in phage binding to bacteria. Therefore, the protein sequence of TFP for each phage is unique, which makes the identification difficult. The alignment of two major TFPs (gp223 and gp37) does not show similarity throughout the whole protein sequence. The variation of the protein sequence among the TFPs is expected because this difference in TFPs contributes to the specific recognition to different types of bacteria. Despite the fact that bxz1 is evolutionarily related to T4, the phage T4 does not infect mycobacteria. Targeting on different hosts could be a result of a distinctive protein sequence, where the C-terminal residues correspond to the recognition and specificity of the receptors from bacteria. Therefore, the diverse protein sequences and structures within TFPs, which are inherent to the protein, makes the identification a challenge.

Demonstration of truncated form of tail fiber protein binding to *M. smeg*

The goal of the study is to demonstrate that gp232 functions as a major TFP in bxz1 that binds to the surface of *M. smeg*. This demonstration of the binding will be inspected by two detection methods, by engineering gp232 with two different tags (the fused-fluorescent tag and the nano-gold tag) for visualization. In general, *Myoviridae* have major TFPs at the distal portion of the tail fibers. Our hypothesis is that the C-terminal domain is sufficient to contribute the binding and specificity. Therefore, gp232 binding to *M. smeg* is demonstrated with a truncated gp232 containing the C-terminal domain, which is a putative binding domain, that is designed based on the secondary structure prediction. The truncated gp232 was cloned from genomic DNA of

mycobacteriophage bxz1 obtained from the Hartful lab, which contains 237 amino acids from the C-terminal region. For fluorescent detection, the truncated gp232 was fused with a green fluorescent protein (GFP) tag at the N-terminus that resulted in a polypeptide chain of 52 kDa.

The fluorescent detection involves attaching a fluorophore at the N-terminal of the protein to assist in capturing the association between the truncated gp232 and *M. smeg* under a fluorescent microscope. This constructed protein (GFP-trunc-gp232) should show affinity and specificity to *M. smeg.* Consequently, the GFP signal should localize at the cell wall of *M. smeg*. The truncated protein was expressed in *E. coli* and purified on a Ni-column followed by gel filtration column. Once pure protein was obtained, incubating the protein with the bacteria clearly exhibited that the *M. smeg* carried the green fluorescence protein. Examining the protein's (GFP-trunc-gp232) reaction with bacteria such as *E. coli* indicated that the protein specifically binds to the molecules on the surface of *M. smeg*, but it does not bind to the molecules on the surface of *E. coli*. Noticeably, this experiment exhibited that the binding of gp232 was specific in its selection of bacteria (Fig 4.3).

Figure 4.3

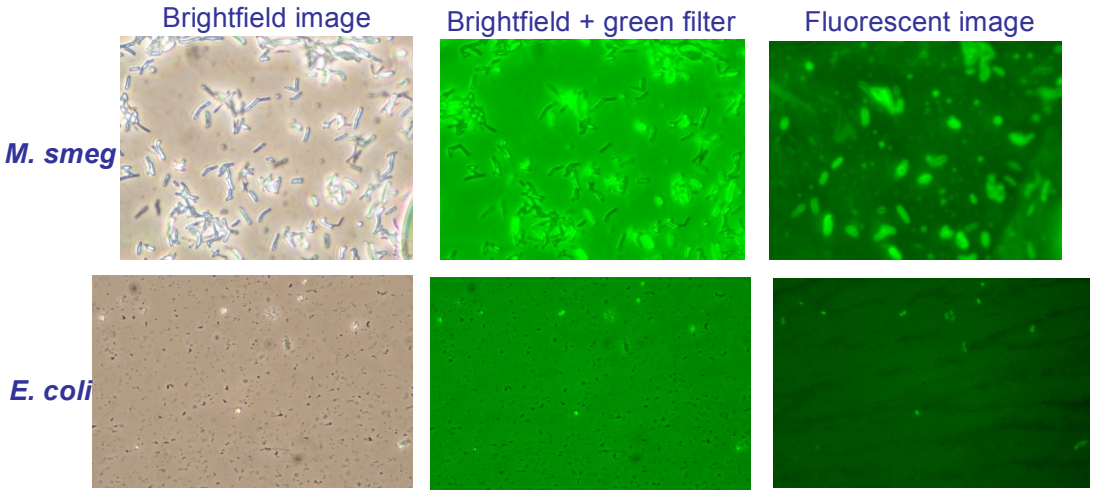


Figure 4.3 Truncated gp232 with an N-terminal GFP tag binds to *M. smegmatis* as observed by fluorescent microscope.

The second detection method was to use the gold nano-particles (nano-gold) as a signal for the binding of the truncated protein gp232, which can be examined and imaged by the electron microscope. The nano-gold particles are commercially available, and a particle size of 5 nm was selected because the nanometer scale matched the size of the mycobacteriophage (10-15 nm) in relation to the size of *M. smeg* (~2−8 μm). They are spherical nano-gold nanoparticles, which are coated with Ni-NTA (nickel-nitrilotriacetic acid). The His-tag of the truncated gp232 would bind to the Ni from the NTA that is conjugated on the surface of the nano-gold particles on incubation. In addition, the binding between gp232 and *M. smeg* will form during incubation, when the binding interface of gp232 makes contact with the surface of *M. smeg*. To allow the nano-gold particles to catch the bound gp232, the first incubation was made with gp232 and *M. smeg*, so that the protein can localize on the surface of *M. smeg*. Then, the incubation was spun down to separate the excess free gp232 that was not fixed with *M. smeg*. After removing the supernatant containing unbound gp232, the nano-gold particles were added to the pellet to explicitly detect the bound gp232 on the surface of *M. smeg*. The nano-gold localized at the cell surface was observed by electron microscopy (TEM) (Fig 4.4). The result supports our proposed function for gp232 as a major tail fiber protein that binds to the mycobacterial surface to allow the phage to carry out the infection.

Figure 4.4

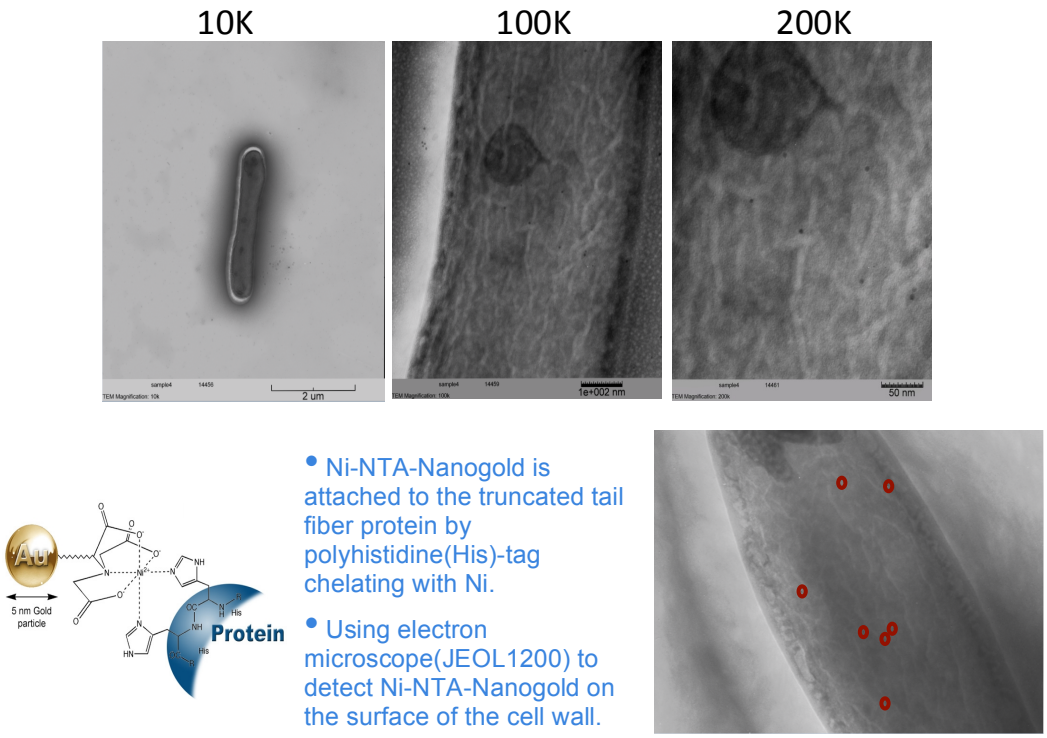


Figure 4.4 Truncated gp232 labeled with Ni-NTA-nanogold binds to the bacterial surface of *M. Smeg* as viewed by electron microscopy (JEOL1200).

The binding of gp232 and the components located in the cell wall of *M. smeg*

The gp232 protein binding on the surface of *M. smeg* indicates that the receptor localizes in the waxy cell wall. If the binding is strong, the truncated gp232 should get pulled down to the cell debris, which contains the cell wall components after cell lysis. To demonstrate the specific binding to the cell wall components from *M. smeg*, the protein was incubated in separate vials with different cell debris, including that from *E. coli*, *Mtb*, and *M. smeg*. Purified gp232 was added to the cell suspension directly before lysis. The results exhibited that the majority of the gp232 (~90%) was pulled down with the cell debris of *M. smeg* (Fig 4.5). This suggests that the binding between the receptor and gp232 is tight enough so that gp232 gets sedimented with the cell wall components, instead of the gp232-receptor complex remaining in the solution. This result implies that the receptor possibly has a moiety that contributes to the strong association to the cell wall. As controls, when the same procedure was applied to *E. coli* and *Mtb* lysates, the protein stayed in the supernatant and was not pulled-down by the cell wall components. This result confirmed the target specificity, in which gp232 recognizes a particular molecule from *M. smeg*.

Figure 4.5

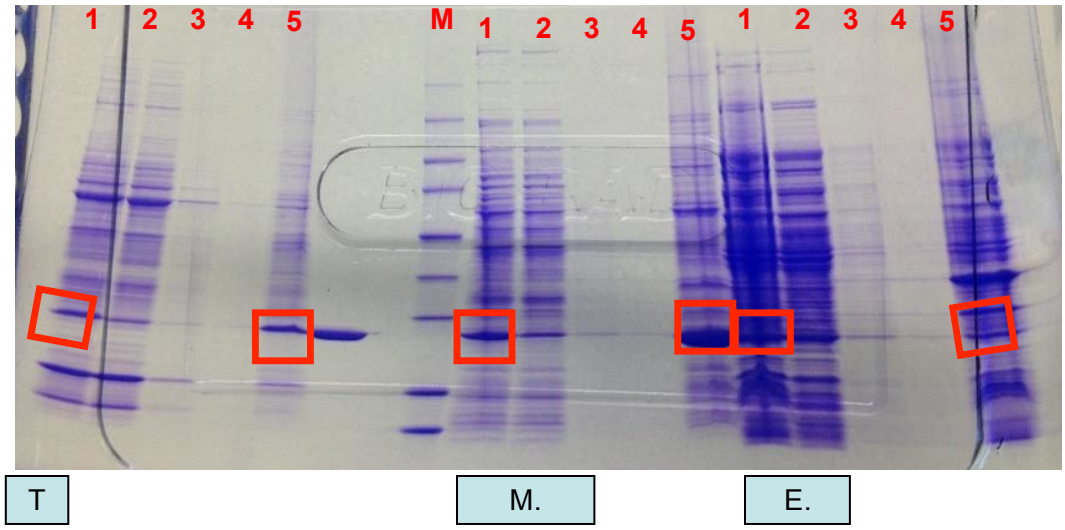


Figure 4.5 Truncated gp232 binds to the cell debris of *M. smeg*. Comparing the amount of gp232 in lane 1, lane 2, and lane 5, the cell debris of *M. smeg* pulled down the most truncated gp232.

Overall structure of the truncated gp232

The truncated gp232 with 237 amino acids (residues 914–1151) and a His tag was purified and the protein was crystallized. For phasing, crystals also were produced with Se-Met incorporated in the peptide chain due to the lack of homolog structure. The best quality crystals for both were obtained using PEG2000 in buffer bicine pH 9.0 with additives (either mannose, NAG, or LDAO). The protein crystallized in the P212121 space group with a dimer of trimers per asymmetrical unit. The unit cell dimensions were a = 90.56 Å, b = 92.85 Å, c = 192.73 Å, α = β = γ = 90.0°. The crystals were sensitive to the condition change in solution when transferring to cryoprotectant (1 M mannose and 25% PEG2000). Data were collected using the synchrotron source (APS and ALS) to 2.0 Å. The structure was solved by the MAD phasing method using the anomalous signal of Se that diffracted to 2.8 Å. The final model consisted of residues from 917–1151 and had a crystallographic R-factor of 0.24 and R-free of 0.30. Nine residues from the N-terminal were not visible in the electron density. Seven of these residues included the initial methionine and six polyhistidines. The structure was built with 234 amino acids from residues 917–1151 and revealed a novel structure of the phage tail fiber protein that had no sequence similarity to any known structure.

It is known that TFPs form a homotrimer, and these three monomers intertwine into an entity as a functional molecule, which also is observed in the structure of the truncated gp232 (Fig 4.6). The crystal structure shows that the overall dimensions of a homotrimer are 110 × 48 × 48 Å, where three polypeptide chains display identical structures (RMSD values of 2.4 Å for C-α atoms). These three subunits of the

homotrimer are twisted and intertwined in a prolonged spiral shape along the longest axis of its overall structure, and the trimer formation results as a 3-fold axis at the center of this trimer. These three polypeptide chains fold into a functional molecule, which shows three domains (head domain, shaft domain, and foot domain). The head domain (residues 917–1039) and the foot domain (residues 1081–1151) are globular and connected by the neck domain (residues 1040–1080) (Fig 4.6). Due to the polypeptide chain packing conformation at the neck domain, the twisting is observed in each polypeptide chain from the first N-terminal domain to the C-terminal domain rotated about 240° along the 3-fold axis. Within the homotrimer, the main force contributing to trimer formation is hydrophobic interactions, which are formed at the central 3-fold axis of the trimer throughout the whole functional unit. One of the signatures is that the hydrophobic residues, such as phenylalanine, from each monomer form a triad geometry at the central axis.

Figure 4.6

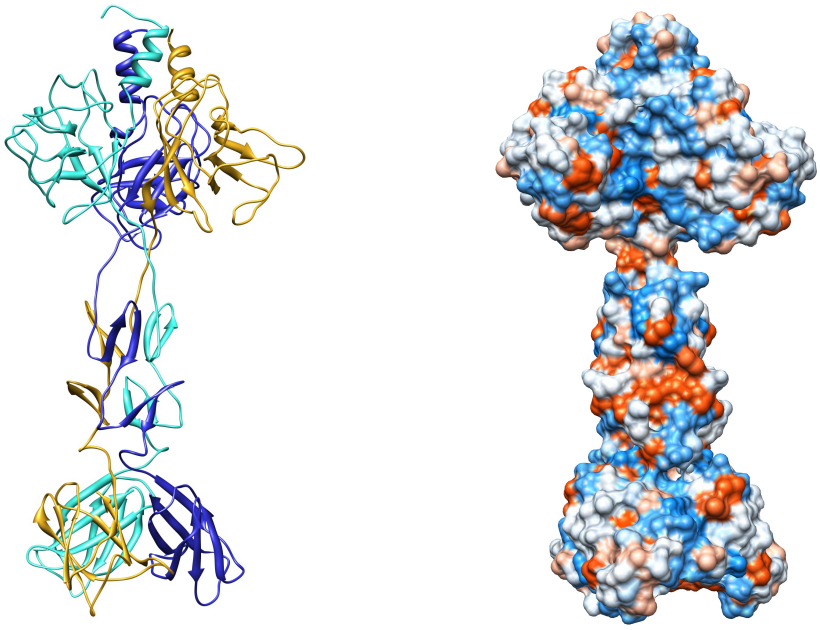


Figure 4.6 Overall structure of the truncated gp232.

The connection head domain

The head domain of gp232 protein, composed of 123 amino acids, is a connection domain to the rest of TFP (residues 1–914) from the N-terminus, and it is the largest domain in the structure of the trunc-gp232 structure (Fig 4.7). The overall secondary structure in the head domain is mainly made of coils, and it has 73% residues in the form of coils. The trimer has a parallel helix bundle (residues 921–928) that is the only helix throughout the whole structure, and it is connected to the rest of the N-terminal residues. Next to the helix, in the head domain, there is a beta globular sandwich domain, which is composed of two beta-sheets (beta 2–4 and beta 5–7). The inner beta sheet is anti-parallel to a beta-meander motif composed of beta 5 (residues 1002–1005), beta 6 (residues 1014–1017), and beta 7 (residues 1032–1037). This beta sheet (beta 5–7) from each monomer is close to the interface of the trimer in the center. There are two long omega-loop motifs connecting beta 5 to beta 6 and beta 6 to beta 7. At the interface, this inner beta sheet utilizes hydrophobic residues, including Leu1036, Ile(1015 and 1035), Val(1003, 1004, and 1005), and Ala1002 to maintain the trimeric state through the formation of hydrophobic interactions. The outer beta sheet is composed of beta 2 (residues 938–940), beta 3 (residues 963–966), and beta 4 (residues 976–979). The residue Phe933 on the first loop connecting the helix to the outer beta sheet (beta 2–4) holds one crucial interaction at the trimer interface. This inter-chain π-π interaction is repeatedly observed at the interface throughout the homotrimer in this structure. Three outer sheets from each monomer form individually into a mixed crossover beta sheet of a psi-loop motif. The overall outlook of this N-terminal head

domain is like a propeller with non-crystallographic 3-fold symmetry along the longest axis. The axis is parallel to the helix bundle and the outer sheets look like blades of a propeller.

Figure 4.7

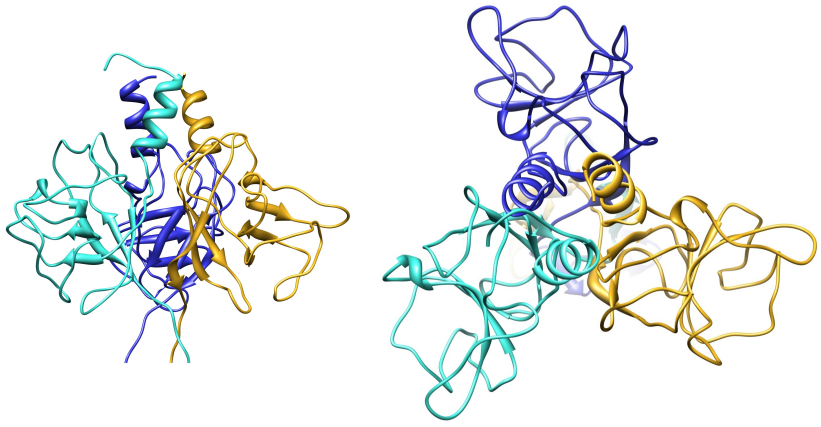


Figure 4.7 Overall outlook of the connection head domain in the truncated gp232. (a) Viewing the connection head domain from the side; (b) viewing the connection head domain along the 3-fold axis from the top.

The shaft domain

This narrow tube-like domain in the structure of this truncated gp232, which contains 42 amino acids in each monomer (residues 133–174), connects the head and the foot domains (Fig 4.8). It is constituted by three highly intertwined subunits, and the trimer twists to an angle of 240°. Each chain has two strand-loop-strand hairpin motifs.

Here, the hairpin motifs from each monomer form a triple beta-spiral fold with the other two monomers. This domain architecture has not been reported in any other tail fiber protein from bacteriophages. However, this beta-spiral fold was found in a few of the fiber proteins in mammalian reovirus and adenovirus.

The long loop (28 Å) connecting the first beta-hairpin motif to the N-terminal connection head domain is the narrowest part of the whole structure (only 10 Å × 10 Å in width), and contributes hydrophobic and hydrophilic interactions to the neighboring chain that sustain the trimeric conformation. The hydrophilic interactions are seen at the interface of the two chains. The oxygen of Asp1056 forms a hydrogen bond (2.8 Å) with the hydroxyl of Thr1045 from the adjacent monomer. Another hydrogen bond at this interface at a distance of 3.1 Å is between the oxygen of Asp1056 and the oxygen of Ser1047. The hydrophobic interactions are made by four sets of hydrophobic residues, which individually form triad geometry, including Val1041, Phe1044, Val1050, and Phe1057. Out of three hydrophobic Phe triad interactions observed in this structure, there are two located in the first layer of the triple beta-spiral (residues 1045–1059). In the second layer of the triple beta-spiral (residues 1060–1076), the force that holds the trimer together is from three hydrophobic triad interactions (Val1067, Leu1074, and Leu1076), and the hydrophobic interactions with neighboring chains. For example, the interaction is found between the **γ**-carbon of Val1061 and the **δ**-carbon Leu1074 of the other chain, and the distance is 3.9 Å. The other hydrophobic interaction is between Val1062 to Val1070 at a distance of 3.5 Å between the **γ**-carbons of Val1062 and Val1070.

Figure 4.8

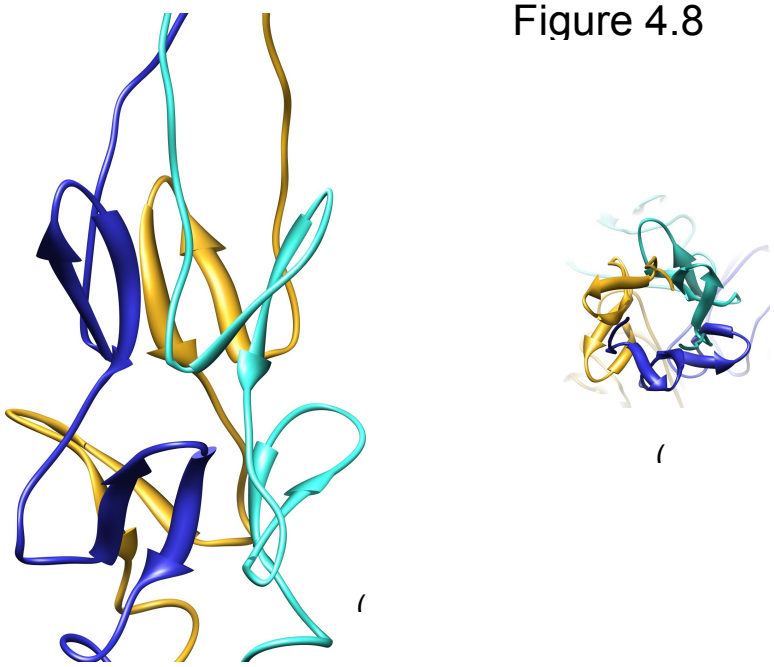


Figure 4.8 Shaft domain of the truncated gp232. (a) Viewing the shaft domain from the side; (b) viewing the shaft domain along the 3-fold axis from the top.

The receptor-recognition foot domain

The C-terminal receptor-recognition foot domain is the tip of the gp232 tail fiber protein, composed of three beta-sandwich motifs. Each beta-sandwich motif has 73 amino acids (residues 1082–1151) and folds into two crossover anti-parallel beta-sheets (Fig 4.9). This motif in the gp232 is different from the greek-key motif found in the foot domain of the other TFPs from *Siphoviridae*. Three beta sandwich motifs are folded and packed into this foot domain with 3-fold symmetry, which is a common feature in most TFPs. These three globular motifs expand 6 Å outward in diameter from the shaft domain, which is the narrowest region of the overall structure, creating a large cavity at

the center of this foot domain. The surface area used to make contact on the bacterial receptors is about 800 $Å^2$, which is much larger than that of gp37 of T4, but it is similar in size with that of the other known TFPs from *Siphoviridae* and *Podoviridae*.

The total number of interactions at the interface between the monomers to maintain the trimeric form in the foot domain is found to be fewer than the other two domains. There are two sets of hydrophobic triad interactions contributed by two residues from each chain (Ala1082 and Val1151). Only one inter-chain hydrophilic contact is found at the interface of the two beta-sandwich motifs, which is the hydrogen bond formed between the oxygen of Thr1104 and the ζ-nitrogen of Lys1084 from the neighboring chain at the distance is 3.1 Å. There are fewer interactions between the dimer found in the foot domain. Therefore, the interactions between these three beta-sandwich motifs from each chain are not as tight as those that are formed in the N-terminal head and the shaft domains. The interactions that hold this foot domain into a trimer appeared only in the upper half of this domain, which is closer to the shaft domain. This suggests that the lower part of the foot domain could be flexible. The lower part of the foot domain, which is the major site for recognizing the receptor from the surface of bacteria, has a cavity that is constructed by three chains at the center. The centered cavity and the loose packing at the lower part of the foot domain suggests that TFP could be flexible at the recognition site. Consequently, this cavity could have the ability to change conformation, or expand to accommodate the receptor that could facilitate the receptor binding.

Figure 4.9

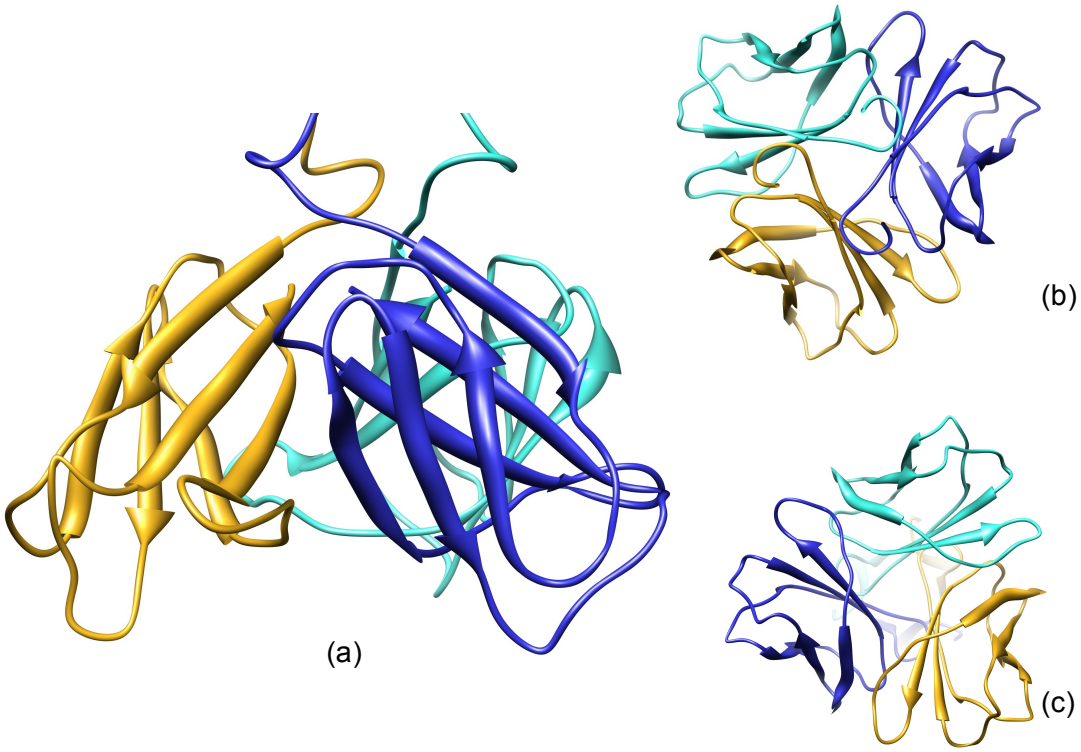


Figure 4.9 Foot domain of the truncated gp232. (a) Viewing the foot domain from the side; (b) viewing the foot domain along the 3-fold axis from the top; (c) viewing the foot domain along the 3-fold axis from the bottom.

Overall structural comparison to bacteriophage tail fiber proteins from *Myoviridae* (T4), *Siphoviridae* (*Lactococcus* phages), and *Podoviridae* (T7)

Due to the lack of the studies on TFPs in mycobacteriophages, the binding mechanism remains unclear. Hence, all of the tail fiber proteins in mycobacteriophages were identified putatively based on sequence alignment. However, the sequence similarity is not sufficient among all the bacteriophage TFPs for unambiguous gene identification. Without sequence similarity to those known TFPs, the TFP identification is difficult. This is a challenge that was faced during the identification of the major TFP in bxz1. Here, the major TFP of mycobacteriophage bxz1 is identified successfully and its structure is the first reported from mycobacteriophage. Currently, there is at least one solved TFP structure for each subclass of dsDNA bacteriophage. Nevertheless, the protein sequence alignment showed that the gp232 from bxz1 has no detectable homology to any of the tail fiber proteins with known structures. Despite the diversity in protein sequences, distinctive structures of TFPs are still evolutionarily related to the function of TFP. Comparing the protein structures available in the database indicates that all TFPs have the intertwined homotrimer conformation equipped with a foot domain at the C-terminus, which is likely used for making contact to bacterial receptors. The structure of tail fiber protein gp232 maintains those features, but it also shows a very unique fold in comparison to the known tail fiber structures.

The structural comparison of gp232 to TFPs from T4

T4 and bxz1 are both categorized under the family *Myoviridae*. The common feature of having two cell-attachment organelles, a minor tail fiber and a major tail fiber, were found in both T4 and bxz1. Despite a common origin between bxz1 and T4, the structures of the major and minor TFPs in T4 have been reported and the overall structures are completely different from gp232 (Fig 4.10). Nevertheless, there is an N-terminal domain of the minor TFP in T4 that is similar to the N-terminal connection head domain in gp232. However, the rest of the protein structure at the C-terminus is formed mainly by coils and does not have any distinguishable domains. The other TFP in T4, gp37, exhibits a very long sword-like structure that is completely different from the structure of gp232. The driving force used to hold three peptide chains into a trimer is abundantly contributed from the hydrophobic triad interactions in the gp232. By contrast, it is the coordination bonds to the 7 iron ions that hold the trimer in gp37 together. Each iron ion forms 6 coordination bonds to the histidines at the center axis of the protein, where the hydrophobic triads are located in gp232. There is no domain formed at the C-terminus of two TFPs in T4, and the last 80 residues on the C-terminus are not as crucial as those in gp232. The last 80 residues in gp232 form a beta-sandwich motif, where three of the residues build a foot domain and present a receptor recognition site that is larger than the major TFP in T4. The significant structural difference that is observed in the TFPs of the T4 and the bxz1 is an indicator of the difference in receptor recognition. It further indicates that these TFPs have specificity on different receptors from the surface of the bacterial wall.

Figure 4.10

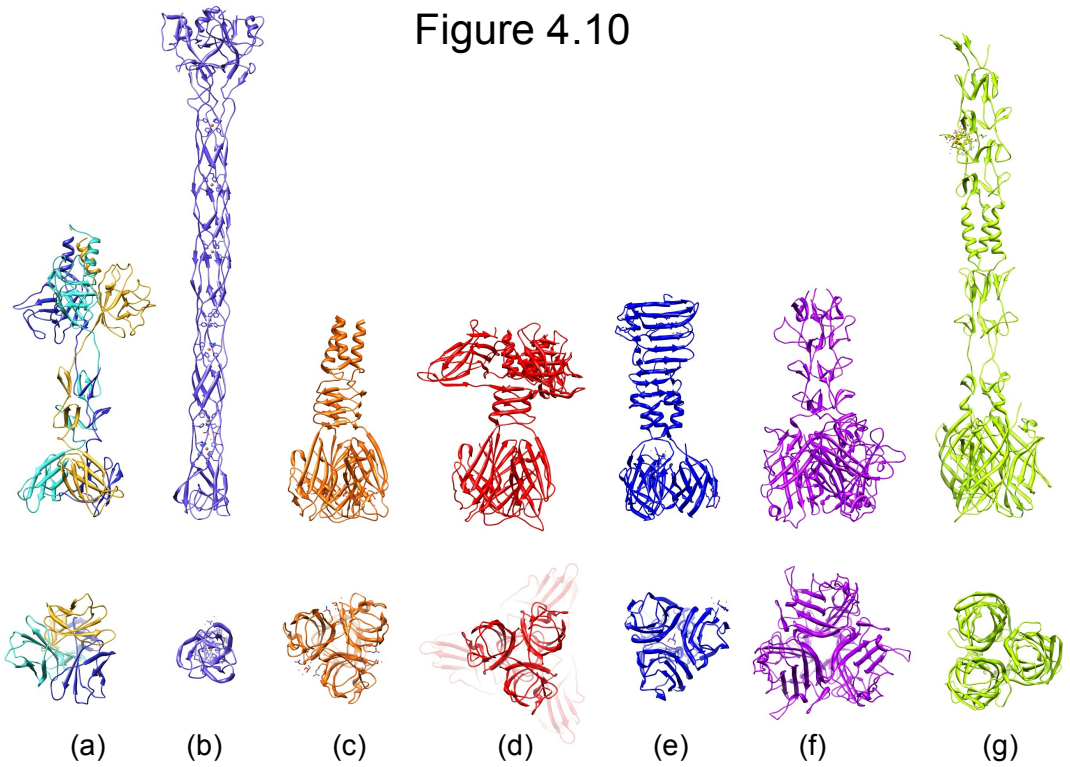


Figure 4.10 Structures of tail fiber proteins from bacteriophages and the viral fiber protein. (a) gp232 from mycobacteriophage bxz1; (b) gp37 (PDB:2XGF) from enterobacteriophage T4; (c) orf49 (PDB:2F0C) from lactococcal phage TP901-1; (d) RBP (PDB: 2BSD) from lactococcal phage P2; (e) gp17 (PDB:4A0T) from bacteriophage T7; (f) fiber protein (PDB:1QIU) from adenovirus; (g) fiber protein (PDB:3S6Z) from reovirus.

The structural comparison of gp232 and TFPs from lactococcal phage in *Siphoviridae*

Siphoviridae has a long non-contractile tail and only the minor tail fibers are attached directly to the base plate. There are only three structures of the minor TFPs that have been solved from the lactococcal phages of the strains TP901-1, P2, and bIL170. All of them share high structural similarity, except for the TFP of P2 because it contains an extra inserted domain (Fig 4.10). The overall structures of these minor TFPs are still formed into a trimer, but are different from the structure of gp232. The two structures from TP901-1 and bIL170 only have a helical bundle at the N-terminus and a triple-stranded beta-helical stalk as the shaft domain. At the end domain at the C-terminus, these minor TFPs have three beta-barrel motifs that form a foot domain. However, the folding of this domain is similar to the foot domain of gp232. Each chain of the TFP in P2 has an extra domain inserted between the helical bundle and the stalk. The three extra domains combined with the helical bundle become a similar assembly as that seen in the N-terminal head domain of gp232. Despite the difference in the shaft domain, the similarity in the domain architecture between gp232 and these minor TFPs from *Siphoviridae* suggests that evolutionary selection might affect the folding of the TFPs. That is why the major TFP in the bxz1 (one of the few *Myoviridae* that infect mycobacteria) has a foot domain similar to the TFPs of *Siphoviridae*, so that it can adapt the efficient phage infection mechanism of *Siphoviridae* to mycobacteria.

The minor TFP of *Podoviridae* T7 shows a foot domain similar to the foot domain of gp232

There is only one solved protein structure of TFP in *Podoviridae* T7, which is one family of the bacteriophage with a substantially shorter tail organelle. The phages in this family lack the long tail seen in *Siphoviridae* and *Myoviridae*, which make them the smallest bacteriophages. It is likely that the short tail of *Podoviridae* is not suitable for infecting bacteria with thick cell walls, which explains why no mycobacteriophage is found in this phage family. Due to the different host range, the TFP of *Podoviridae* T7 has the least probability of having any structural similarity with the TFPs of bxz1. The middle domain of the TFP (gp17) in T7 is a beta-structured pyramid domain. This domain has been seen in the proteins that make the assembly of the cell-penetrating device in *Myoviridae* (phage P2 and phi92), but it is very different from the shaft domain of gp232 (Fig 4.10). However, the structure of this gp17 in T7 shows that it has a foot domain with a size and fold that are unexpectedly close to what is seen in the foot domain of the gp232. Currently, gp17 is the only known TFP to have a foot domain similar to that in gp232. This indicates that the overall fold of the foot domain does not define the specificity for recognition, but the exact residue composition or the electrostatic potential on that recognition surface defines the host specificity.

The potential planar recognition site at the foot domain of gp232

The foot domain, which is built from only the C-terminal residues of the monomers, results in a tetrahedron conformation with four triangular faces. Three of the

triangular faces, connecting the shaft domain, are the interfaces of any two monomers from the homotrimer. The last triangular face is the only face of this foot domain containing all three monomers of the trimer. It is located at the bottom, which implies a potential phage-bacteria recognition site. This planar recognition site of gp232 is perpendicular to the three-fold symmetry axis of the homotrimer (Fig 4.11a). It may be speculated that this site is where mycobacteriophage bxz1 makes the first contact with the mycobacterial surface. At the center of the recognition site, there is an entrance to a cavity that is located at the center of this foot domain. It is known that tail fiber plays an essential role at the initial stage of phage infection by making contact with the bacterial surface. The tip of the tail fiber that is composed of the C-terminal residues of TFP provides the attachment and the host specificity. This has been demonstrated in tuning the infection of phage P2 by swapping the C-terminal residues with the other TFPs. Therefore, the specificity of bxz1 targeting on *M. smeg* would likely come from the recognition site and the cavity in the foot domain of the major TFP.

Figure4.11

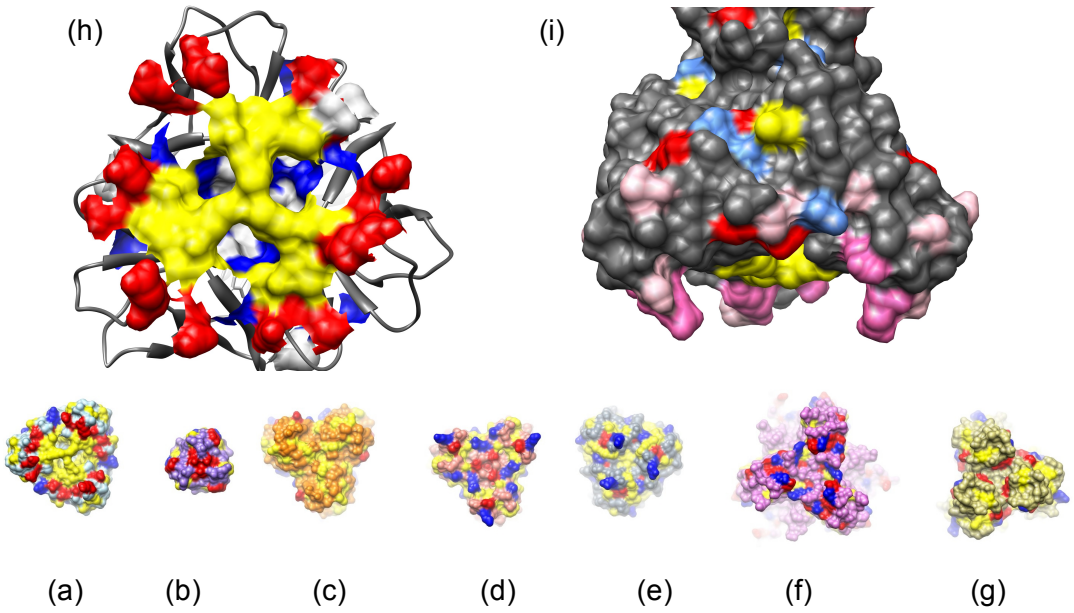


Figure 4.11 Potential receptor-recognition surface of the bacteriophage TFPs and the vial fiber proteins. (a) gp232 from mycobacteriophage bxz1; (b) gp37 (PDB:2XGF) from enterobacteriophage T4; (c) orf49 (PDB:2F0C) from lactococcal phage TP901-1; (d) RBP (PDB:2BSD) from lactococcal phage P2; (e) gp17 (PDB: 4A0T) from bacteriophage T7; (f) fiber protein (PDB:1QIU) from adenovirus; (g) fiber protein (PDB:3S6Z) from reovirus; (h) the bottom view of the cavity with three different surface potentials, which is centered at the receptor-recognition site [red, negatively charged residues (Asp and Glu), blue; positively charged residues (Arg and Lys); yellow, hydrophobic residues (Ala, Val, Leu, and Ile)]; (i) side view.

The structural insights of the foot domain, including the recognition site and the cavity, could suggest the phage-bacteria recognition mechanism. The structures of the

foot domain in the TFPs are likely to be diverse to accommodate different bacterial receptors on the surface. Each TFP might have a different structure at this recognition-binding site to adapt its distinctive and selective interaction with the receptors on the bacterial surface. The recognition binding site of gp232, containing 12 Ser (residues 1116,1133,1138, and 1142), 6 Asp (residues 1113 and 1118), and 6 Glu (residues 1135 and 1137) in the trimer, is a potential binding site with a dense population of negatively charged and hydrophilic residues. These residues are lined up at the edge of the triangular binding face, displaying a negatively charged patch of two Asp residues and two Glu residues on each side. This patch could make the initial contact by interacting with the positively charged atoms on the surface of the bacteria, which may lead to further receptor recognition binding to the other parts of the foot domain. The additional interactions could be contributed by the cavity formed at the center of the foot domain. It has an entrance at the center of the binding site with three hydrophobic Leu residues guarding the entrance. The property of the cavity is important because it might assist in determining the specificity and tight binding to a receptor. The cavity is in a mushroom-like shape that has a hydrophobic stem and a hydrophilic cap region (Fig 4.12). The stem of this mushroom-like cavity serves as a portal and is formed by hydrophobic residues 1119–1122 (AVLV), Phe111, and Trp1124. One of the residues, Leu1121, is on the stem where it connects to the recognition site. On the other side of the stem is the deepest part of the cavity, which is the cap of the mushroom. This cap is hydrophilic and composed of 3 His1108 and 3 Ser1110. It is possible that, on binding, the structural conformation changes to open the narrowest part of hydrophobic stem at Leu1121, and

the hydrophobic residues of the stem clamp on the hydrophobic part of the bacterial receptor. Moreover, the region inside the cavity between the stem and cap contains three small pockets. The residues (R1112, R1146, W1124, T1148, and S1110) of each monomer each construct a small compartment with positive potential on the surface. The structure of this foot domain shows that, from the recognition site to the cavity inside, the potential on the surface is from negative charge to hydrophobic at the stem region and to positive charge at the cap region. Therefore, the negatively charged residues (Asp1118, Asp1113, Glu1135, and Glu1137) at the recognition site are used to grab onto the cell surface. On contact with the bacterial surface, the cavity could form interactions with a specific receptor.

Figure 4.12

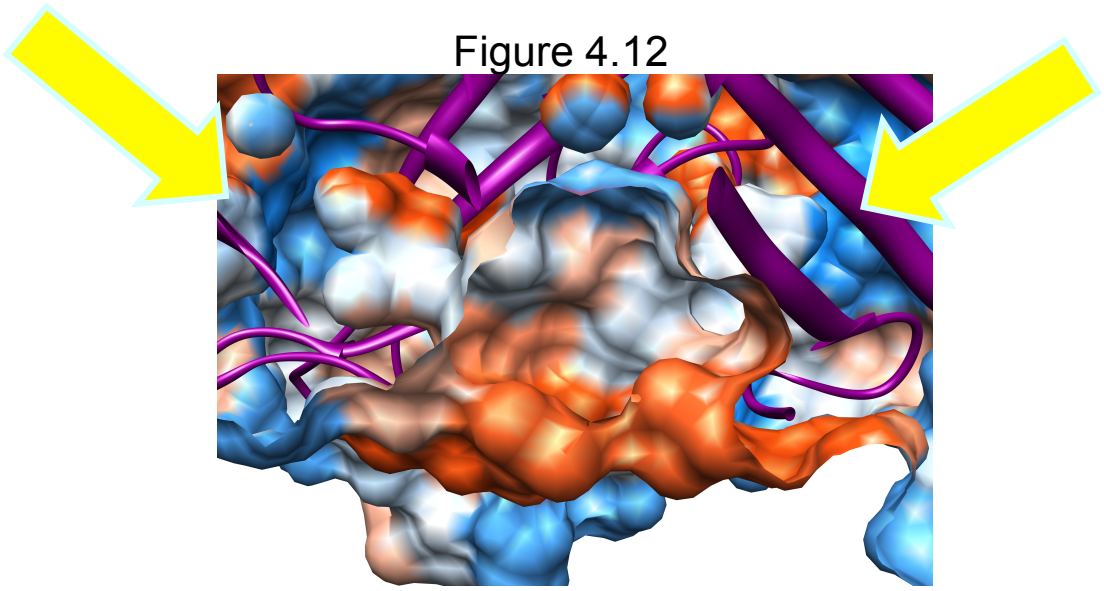


Figure 4.12 A mushroom-like cavity at the center of potential receptor-recognition surface. The surface is colored by hydrophobicity (orange: hydrophobic, blue: hydrophilic).

Structural comparison of the potential receptor-recognition site among all the bacteriophage TFPs

The overall structure of the truncated gp232 shows that it is unique among the known structures of other TFPs in bacteriophage. Apart from the structural diversity among all TFPs, the residues at the recognition site also are the products of co-evolution between phages and bacteria. Therefore, the residues for making interactions at the recognition site are specific toward the distinct cell wall components they interact with. A comparison of the recognition sites of all known TFPs shows that the interactions between the bacterial receptor and the recognition site can be classified into three types. These interactions are made either by positively charged, hydrophobic, or negatively charged residues from the TFPs (Fig 4.11a-g). The first type of interaction is through positively charged residues and has been shown in the minor TFP (gp12) from phage T4. This minor TFP assists the base plate at the end of the phage tail to dock on the bacterial cell wall. The residues, Arg and Lys, at the tip of the receptor-binding site of gp12 are responsible for the interactions with the bacterial surface. Another TFP from the lactococcus phage P2 and the viral fiber protein also have positive residues, Arg and Lys, at the recognition site, which make the initial contact. The second type is primarily made by hydrophobic residues, which might be Trp, Phe, Leu, Val, or Ala at the binding site. The TFP of the lactococcus phage TP901-1 presents a large hydrophobic area, in which the receptor-binding site contains Trp, Val, and Ala. A negatively charged residue such as Tyr at the recognition site is seen in the structure of gp37 from T4 phage. Among all the known structures of TFPs, gp17 from the T7 phage contains a similar

architecture in the foot domain, with a cavity that resembles the foot domain of gp232. However, there are a few differences. Unlike the Asp and Glu at the recognition site of gp232, the first layer of the recognition site in gp17 contains six positively charged Arg residues. Three of the six Arg residues point outward from the protein and are potentially responsible for making the first contact on the bacterial surface. This recognition site, similar to that in gp232, has a planar face with an entrance to a cavity at the center. Unlike the narrow portal observed in gp232, the cavity entrance of gp17 is completely blocked by a hydrophobic residue (Ala518). The residues found at the entrance in gp17 are hydrophobic, which resembles the hydrophobic portal in gp232. At the entrance of gp17, the residues Ala518 and Asp520 have been reported to hold the binding specificity[133]. This implies that the Leu1121 in gp232, which partially blocks the portal, may play a role in binding specificity. The cavity of gp17 is smaller, with an absence of a positively charged pocket. Thus, the potential binding cavity of gp17 has only hydrophobic residues (Phe and Ile) formed inside and positively charged residues at the recognition site. Therefore, gp232 is the only TFP that has a cavity right next to the entrance that is not completely blocked by hydrophobic residues.

The structural comparison to viral fiber proteins

The overall structure of truncated gp232 is novel, and the shaft domain has a unique fold that has not been seen in any of the tail fiber proteins from bacteriophages. However, the shaft domain resembles the triple beta-spiral motif reported in the fiber protein of avian reovirus and adenovirus. Similar to that in gp232, the structures of the

viral fiber protein contain a similar and long triple beta-spiral motif in its shaft domain. The domain architecture of these viral fiber proteins is the same as that of gp232. They also contain three domains known as the tail, the shaft, and the knob domain from the N-terminus to the C-terminus, which may account for the head, the shaft, and the foot domain in gp232, respectively. These viral fiber proteins assist the viruses in choosing the hosts in order to induce the infection after fiber proteins attach to the host cell receptors. For example, adenovirus seratype 37 (Ad37) uses the viral fiber protein for initiating viral infection in humans and leads to severe epidemic keratoconjunctivtis (EKC), a contagious ocular disease[134]. This fiber protein has been shown to bind two receptors on the cellular surface of the host, including GD1, a glycan, and the virus receptor from the host immune system. The structure of the other fiber protein from reovirus, named σ1, contains two sets of beta-spiral motifs in the shaft domain, connected by a helices bundle, and the foot domain shares a similar fold as that in the fiber protein of Ad37. The length of the shaft domain in these viral fiber proteins is longer than that in gp232. The shaft domain of gp232 has two beta-spiral repeats and its length is 24 Å, whereas the shaft domain of AD37 has four beta-spiral repeats with a length of 46 Å (PDB:1QIU). The length of the two sets of beta-spiral motifs in σ1 is 42 Å with four repeats and 57 Å with another three repeats (PDB:3S6Z).

Less is known about the interaction of bacteriophage TFP with receptor on the bacterial cell surface. Interestingly, it has been shown that the cell-surface glycans bind between the second and third repeat of the beta-spiral motif from σ1, instead of the receptor-recognition site at the foot domain. The protein structure of σ1 in complex with

di-sialic acid (PDB:3S6Z) reveals a carbohydrate-binding site. Due to the structural similarity shared between the shaft domains of the σ1 and gp232, the beta-spiral motif in the gp232 might serve as a tool to recognize the carbohydrate moieties on the cell surface of mycobacteria. Compared to that in the σ1 structure, the potential carbohydrate-binding site of gp232 is likely to be at the region composed of the beta-spiral repeat 1 (residues 1043–1059) and Ser1071 from the beta-spiral motif repeat 2.

There are two carbohydrate-binding sites in the foot domain reported in the viral fiber proteins from adenovirus and reovirus. The adenoviral fiber protein shows GD1a-glycan binding at the interface of trimer in the foot domain, which is related to the receptor-recognition site of gp232. However, the residue composition in the foot domain and the recognition site of the AD37 fiber protein (PDB:3N0I) is very different from the planer recognition site in gp232. Residues of Lys, Tyr, Ser, and Thr form the carbohydrate-binding site of AD37 fiber protein, where the electrostatic surface potential is different from that of the Asp-Glu at recognition site of gp232. At the interface of the two monomers in the foot domain of gp232, there is a shallow groove constructed by Tyr1085, Lys1084, Thr1104, and Lys1114, which has similar residue composition with that in AD37, and could serve as a binding site. The location of this shallow groove in gp232 can relate to a carbohydrate-binding site of σ1 composed of residues Asn, Gln, and Thr, where GM2 and GM3 (glycans on the cellular surface) bind (PDB:4GU3 and 4GU4). Based on the structures of the two viral fiber proteins bound to receptors, structural comparison indicates two new potential sites, in addition to the planar recognition site, for binding the glycolipids on the mycobacterial surface.

4.3 Methods

Cloning, expression and purification of the truncated gp232 and the truncated gp232 with GFP-tag

The gene *gp232* was identified as a TFP in mycobacteriophage bxz1, and was amplified in a truncated form with the forward primer with an *Nde*I restriction site and the reverse primer with a *Hind*III restriction site. The template was the gDNA of mycobacteriophage bxz1 that was obtained from the Hartful lab. The PCR product was purified and digested with the restriction enzymes. This double-stranded DNA was ligated into the pET30b vector (Novagen) with kanamycin as antibiotic selection. The plasmid containing truncated *gp232* (pET30b-trunc-gp232) was transformed into *E coli* BL21 cell line (Novagen), and the transformants were plated on Luria-Bertani (LB) agar medium in the presence of kanamycin (50 μg/mL). The other construct of the truncated *gp232* containing a green fluorescent protein (GFP) tag at the N-terminus also was cloned into the pET30 vector. The gene *gfp* was cloned from the plasmid of pAcGFP1 from Clontech with the forward primer with *Nde*I restriction site and the reverse primer with *Bam*HI restriction site. This GFP-tag was engineered into the pET30b vector to make the pET30b-gfp vector. The truncated form of the gene *gp232* (*tgp232*) was cloned again, containing different restriction sites. The PCR product of *tgp232* was digested with the restriction enzymes and ligated into the pET30b-gfp vector to make the pET30b-gfp-tgp232 construct. The product plasmid was transformed into *E. coli* BL21 and selected with kanamycin on LB-agar plates. A single colony of both constructs (pET30b-trunc-gp232 and pET30b-gfp-tgp232) was picked and used to inoculate the

starter culture (20 mL) containing kanamycin (50 μg/mL). The cultures were grown to an OD_{600} of 0.6 at 37°C. Each 20 mL of the inoculant was added to 2 L LB media. After incubating in a shaker about 4–5 hours (OD_{600} ~ 0.6 to 0.8), IPTG was added to the cultures at a final concentration of 1 mM to induce protein overexpression, and the cultures were further incubated in the shaker for 16 hours at 18°C.

The cells were harvested and spun down by centrifuging at 3000×*g* for 40 minutes. The pelleted cells were resuspended in the lysis buffer, containing 50 mM Tris pH 7.5, 500 mM NaCl, and 2 mM β-mercaptoethanol. DNase at a final concentration of 10 μg/mL and protease inhibitor cocktail (Novagen) were added before lysing with a high-sheer fluid processor. The cell resuspension was passed through a chamber with a diameter of ~150 microns under high pressure (18000 psi). The cell debris was separated from the crude cell lysate by centrifuging at 10,000×*g* for 60 minutes at 4°C, and the supernatant contained the soluble protein that was designed and engineered in the *E. coli* plasmid. The clarified lysate was passed through a 0.22-μm syringe filter before the subsequent chromatographic purification steps.

Both protein constructs (tunc-gp232 and GFP-tgp232) were obtained at ~ 95% purity by affinity column and gel filtration column. Due to the similar pI value of both proteins, they were treated with the same purification procedure. The first step was to utilize the affinity of the designed N-terminal His-tag to the immobilizing nickel-affinity column, which pulled out our target proteins from the lysate. The protein was bound to the nickel-affinity column. After washing the column with 5 column-volumes of the buffer A, the protein was eluted out with a gradient elution of buffer B (20 mM Tris pH

7.5, 500 mM NaCl, 2 mM β-mercaptoethanol, and 500 mM imidazole). Based on the absorbance readout at the wavelength of 280 nm recording along with the elution gradient, the bound protein was found to elute out at ~100–150 mM imidazole. Those fractions were pooled together and concentrated to 2 mL using a Centricon concentrator (cutoff 10 K). The protein sample was loaded onto a gel filtration column (Superdex 200 from GE/Pharmacia) and purified to homogeneity in the buffer with lower salt concentration (20 mM Tris pH7.5, 150 mM NaCl, and 2 mM β-mercaptoethanol). Both protein constructs of the truncated gp232 were eluted in a major peak that identified the molecular weight of both samples in that fraction. The results showed both proteins formed into a trimer in solution (~75 kDa for trunc-gp232 and 150 kDa for GFP-tgp232). Those fractions containing homogeneous protein were collected and concentrated to about 6 mg/ml.

Se-Met protein production

The plasmid with inserted gene of the truncated *gp232* was transformed in the *E. coli* BL21 cell line. A 5-mL cell culture was inoculated with a single colony and grown at 37°C for 8 hours. This was added into the starter culture of 250 mL M9 media. The starter culture was incubated in the shaker at 37°C for 16 hours. For the large-scale prep, each liter of the minimal M9 media was inoculated with 20 mL of the starter culture, and the cells were grown at 37°C. When the OD_{600} reached 0.3, the cell culture (1 L) was supplemented with amino acids mix containing 50 mg of isoleucine, leucine, and valine, and 100 mg of lysine, phenylalanine, and threonine. After 30 minutes of shaking, the Se-

methionine (60 mg) was added into the culture. Once the cell density reached OD_{600} of 0.6, IPTG was added to induce protein expression, and the cell cultures were incubated in the shaker at 16°C for a further 18 hours. The cells were harvested by centrifugation. The procedures of lysis and purification were the same as that in the protocol for purifying the native protein. The protein was eluted at 100–200 mM imidazole in the gradient elution. The fractions containing the pure proteins were pooled and dialyzed against the buffer (20 mM Tris pH7.5, 150 mM NaCl, and 2 mM BME). The protein was concentrated to 6 mg/mL for crystallization trials.

Culture of *M. smeg* (mc^2 155), *M. tuberculosis* (mc^2 7000), and *E. coli*

The cell culture of *M. smeg* (mc^2 155) was inoculated in media with 1 µL of the glycerol stock stored at −80°C. Middlebrook 7H9 broth media with supplements of 0.2% dextrose, 0.085% (w/v) NaCl, and 0.05% (v/v) Tween 80 was used. The cells of *M. tuberculosis* also were inoculated in the Middlebrook 7H9 media with an extra supplement of 50 µg/mL pantothenate. The cell culture of *E. coli* (strain BL21) was inoculated in LB broth. Both cell cultures were harvested for the binding demonstration with gp232 after incubating at 37°C in a rotary shaker to reach an OD_{600} of 0.6.

Imaging with GFP-tag protein

To identify the function of gp232 as a tail fiber protein in the mycobacteriophage bxz1, the truncated gp232 was fused with a GFP-tag at the N-terminus and used to demonstrate the binding to the bacteria under a fluorescent microscope. The cultures of

M. smeg and *E. coli*, both grown to 0.6 of OD_{600}, were transferred to two different slides. The protein (GPF-tgp232) also was added on each slide. The bacteria (*M. smeg* and *E. coli*) were incubated with protein on the slide for 60 minutes. The cover slide was used to cover and fix the bacteria on the focal plane. Two samples were examined under the microscope (Nikon, Eclipse TS100-F) equipped with a dichronic mirror in the green filter block.

Electron microscopy

The truncated gp232 was attached with nano-gold (NG) particles to observe the binding between gp232 and the surface of the bacteria using an electron microscope. The NG-particles (5 nm, Nanoprobes Company) were incubated with protein on ice, and a Ni-NTA (nickel-nitrilotriacetic acid) moiety coated on the surface of the spherical nano-gold particles formed coordination bonds to the His-tag of the protein. This complex (NG-trunc-gp232) was applied to a carbon mesh with *M. smeg* and images were taken by the electron microscope (JEOL1200).

Affinity to the bacteria

To compare the affinity for *M. smeg*, *M. tuberculosis*, and *E. coli*, each was harvested from a 5-mL cell culture. The cells were spun down and suspended in 1.5 mL of buffer (20 mM Tris pH 7.5 and 300 mM NaCl). DNase and protease inhibitor cocktails were added, followed by adding protein (trunc-gp232, 6 μM). In the presence of the protein, the cells were lysed with a mini-bead beater (Biospec products). The

lysates were centrifuged at 10,000×*g* for 30 minutes to obtain cell debris and clarified lysate. The cell debris of each was washed with buffer (20 mM Tris pH 7.5 and 500 mM NaCl), and then spun down. The wash step was repeated twice. Samples were taken from each step for SDS-page analysis.

Protein crystallization

The truncated gp232 of the native form and the truncated gp232 with the Se-methionine modification were purified to homogeneity. Both proteins (6 mg/mL) were screened under 768 crystallization conditions using the sitting-drop vapor-diffusion method. Each condition was set up with the volume ratio of 1 to 1 (protein to well solution) by a liquid handling robot (mosquito, TTP) in 96-well sitting-drop plates (Art Robbins Instruments). The crystallization plates were placed in an environment that maintained a constant temperature and limited vibration. Protein crystals appeared in the hit conditions after 48 hours, and the quality of the crystals was optimized under several hit conditions in the 24-well sitting-drop plate. The best crystals were obtained in a crystallization condition containing 100 mM bicine pH 9.0, 20% PEG200, and 4 mM beta-OG.

Data collection and structural determination

The crystals of the truncated gp232 containing the Se-Met modification were sent to the Hung lab and collected in the synchrotron facility of the advanced light source (ALS) at the Lawrence Berkeley National Lab. The crystals were mounted in a loop

under the cryogenic condition that added ethylene glycol at the final concentration of 30% (v/v), and flash frozen in the liquid nitrogen. The crystals diffracted to a resolution of 2.5 Å and 2.8 Å, and 180° of data were collected using 1°-oscillation range at three different wavelengths for the peak, inflection, and remote data.

Three datasets collected at the different wavelengths were indexed using HKL2000, which showed that the protein crystallized in the space group P212121 with the unit cell dimension (peak data) of a = 91.290 Å, b = 93.148 Å, c = 193.652 Å, $\alpha = \beta = \gamma = 90°$. All three data were separately integrated and scaled in HKL2000. The anomalous signal from the Se in the peak and inflection data was used for phasing. The crystal structure was solved by the multi-wavelength anomalous dispersion (MAD) method. The scaled data from the three datasets were analyzed by AutoSharp to calculate the phase. The initial phase information, the X-ray intensity data, and the protein sequence were used as input in AutoSol (PHENIX) that generated an initial structural model of our target protein, the truncated gp232. The initial model was viewed in COOT, and was fixed to fit in this initial electron density. The final structure was obtained after few cycle of refinement (CCP4 and PHENIX).

5. CONCLUSION

This study of lipid-binding proteins focuses on two deadly pathogenic diseases, tuberculosis and malaria, which have caused infections that threaten more than one-third of the human population. Although both diseases currently have drugs that can cure the infection, the emergence of drug-resistant strains reveals the urgent need to find new treatments. To eliminate the pathogens, the insight gained from the structures of lipid-binding proteins can facilitate our search to find inhibitors for the drug target proteins, understand the pathogen physiology, and design novel targeted drug delivery. To achieve these three goals, the structural studies of lipid-binding proteins were performed to identify the protein functions that can be used to understand pathogen physiology and for designing therapeutics, and to explore the inhibitor-binding site of a drug-target protein. The crystal structure of each lipid-binding protein (LprG, LprA, PfENR, and gp232) provided knowledge that can be applied to the goal for each protein, which is set to approach pathogen eradication from a specific angle.

LprG and LprA, two proteins from Mtb, have been reported to be capable of inducing the host immune response by activating TLR2. LprG is essential for Mtb to survive during infection. The structural determination of LprG and LprA has connected the knowledge of genomics and pathogen physiology. The ligand identification study showed that LprG binds precursors of LAM and LM, such as tri-acylated glycolipid (Ac1PIM2). LprG also can bind LAM and LM. This indicates that the function of LprG

is a glycolipid carrier. Previous studies that showed that LprG is on the same operon with Rv1410c, a sugar efflux pump, and that the function of Rv1410c requires LprG. The result that LprG is a glycolipid-binding protein suggests that LprG has a potential to mediate sugar transport. Therefore, the function of LprG is likely associated with glycolipid biosynthesis and maintenance of the integrity of the cell wall. The mutant LprG (V91W), which was made based on the structural guidance, verifies that LprG binding to tri-acylated glycolipids induces the TLR2 reaction. The crystal structure of LprA revealed a binding with a di-acyl lipid. The ligand identification showed that LprA binds diverse forms of phospholipids (PG, PE, PI, and PIP). This suggests that LprA is a phospholipid-binding protein. In combination with the genomic data, these results suggest that LprA is on the same operon with Rv1272c and Rv1273c, which are homologous to the phospholipids/ lipopolysaccharides transporter (MsbA) in *E. coli*. The result of LprA binding to phospholipids suggests that LprA, like LprG, might associate with membrane proteins (Rv1272c and Rv1273c) in assisting phospholipid transport. Here, the structural studies of LprG and LprA give the first important evidence for identifying the protein function. This provides an example of using protein structure as a guide in linking the protein function to pathogen physiology. Ideally, the collected information is very useful in understanding the mechanism of pathogen survival that would lead to the eradication of the pathogen.

The enoyl reductase in the type II fatty acid biosynthesis (FAS II) pathway is a validated drug target protein in treating pathogenic diseases like malaria because humans have only the type I fatty acid biosynthesis (FAS I). In malaria, the FAS II pathway is

essential when a parasite (*Plasmodium falciparum*) enters hepatocytes and multiplies rapidly. The functional mechanism of a known anti-microbial agent, Triclosan, has been demonstrated; by inhibiting enoyl reductase (PfENR) and blocking the FAS II biosynthesis pathway. The IC_{50} value of 70 nM indicates that it is a potent inhibitor of PfENR *in vitro*. However, an enzyme in the human liver could glucuronate triclosan and disrupts the hydrogen bonding to Tyr277, which weakens the affinity and decreases the efficacy.

The first step is to generate the structure-activity relationships (SAR) by screening hundreds of triclosan analogs that would help in finding the key interaction between protein and inhibitor. The second step is to apply structural insight of those key interactions along the triclosan scaffold onto new sets of chemical scaffolds, which will have high affinity and sustain the potency in liver. The SAR study of triclosan-like analogs suggests that modification at the 5-position on ring A can maintain the same potency as triclosan in the enzyme assay. The modification at the 5-position reveals a novel hydrophobic binding site that is observed on Phe368 flipping. This site could be used to increase affinity without reducing inhibition, because the 5-fluoro-phenyl analog showed a 2-fold stronger inhibition to triclosan. The structural knowledge that we gained in the SAR study from analyzing triclosan-like analogs assisted us to obtain the most potent inhibitors (**698** and **325**) against PfENR, with a core scaffold of tetrahydroquinoline and piperidine. In finding potent inhibitors, the structural study of PfENR helped us to generate prominent inhibitors by fine-tuning the modifications on

the core scaffold. This provides the possibility to improve the pharmacokinetic issue that triclosan has encountered.

Mycobacteriophages use tail fiber proteins (TFPs) to make contact on the bacterial cell wall, and the binding between phage and bacteria is tight and specific. The protein identification and the structural study of TFPs from mycobacteriophage bxz1 were performed to understand this binding mechanism, which can be applied in a targeted drug delivery system. The protein sequence analysis indicated that TFP in mycobacteriophage bxz1 is gp232, and the crystal structure is the first solved among all mycobacteriophages. The role of gp232 as a TFP in bxz1 has been characterized by observing the specific and tight binding between gp232 and *M. smegmatis*. Here, the protein structure helped us to identify the TFP in bxz1, and guided us in designing the protein conjugation. On conjugating gp232, the binding specificity and affinity to mycobacteria could allow the targeted drug delivery to reduce drug toxicity and to enhance drug efficacy by increasing the local concentration at the infection site.

REFERENCES

1. Oresic, M., *Metabolomics, a novel tool for studies of nutrition, metabolism and lipid dysfunction.* Nutr Metab Cardiovasc Dis, 2009. **19**(11): p. 816-24.

2. Carpenter, K.J., *Early ideas on the nutritional significance of lipids.* J Nutr, 1998. **128**(2 Suppl): p. 423S-426S.

3. Katan, M.B., P.L. Zock, and R.P. Mensink, *Effects of fats and fatty acids on blood lipids in humans: an overview.* Am J Clin Nutr, 1994. **60**(6 Suppl): p. 1017S-1022S.

4. Zhang, Y.M. and C.O. Rock, *Membrane lipid homeostasis in bacteria.* Nat Rev Microbiol, 2008. **6**(3): p. 222-33.

5. Shores, D.R., et al., *New insights into the role of fatty acids in the pathogenesis and resolution of inflammatory bowel disease.* Inflamm Bowel Dis, 2011. **17**(10): p. 2192-204.

6. Guenin-Mace, L., R. Simeone, and C. Demangel, *Lipids of pathogenic Mycobacteria: contributions to virulence and host immune suppression.* Transbound Emerg Dis, 2009. **56**(6-7): p. 255-68.

7. Vilcheze, C. and W.R. Jacobs, Jr., *The mechanism of isoniazid killing: clarity through the scope of genetics.* Annu Rev Microbiol, 2007. **61**: p. 35-50.

8. Shi, L., et al., *Carbon flux rerouting during Mycobacterium tuberculosis growth arrest.* Mol Microbiol, 2010. **78**(5): p. 1199-215.

9. Daniel, J., et al., *Induction of a novel class of diacylglycerol acyltransferases and triacylglycerol accumulation in Mycobacterium tuberculosis as it goes into a dormancy-like state in culture.* J Bacteriol, 2004. **186**(15): p. 5017-30.

10. Alvarez, H.M. and A. Steinbuchel, *Triacylglycerols in prokaryotic microorganisms.* Appl Microbiol Biotechnol, 2002. **60**(4): p. 367-76.

11. Paul Crellin, C.-Y.L., Yasu Morita, *Chapter 6 Metabolism of Plasma Membrane Lipids in Mycobacteria and Corynebacteria in book (Lipid Metabolism)*. USA: InTech.

12. Mishra, K.C., et al., *Functional role of the PE domain and immunogenicity of the Mycobacterium tuberculosis triacylglycerol hydrolase LipY.* Infect Immun, 2008. **76**(1): p. 127-40.

13. Ortalo-Magne, A., et al., *Identification of the surface-exposed lipids on the cell envelopes of Mycobacterium tuberculosis and other mycobacterial species*, in *J Bacteriol*1996. p. 456-61.

14. http://abbys-humanphysiology.wikispaces.com, *Triacylglycerol figure*, TGAFig, Editor.

15. Cooper, G.M., *The Cell: A Molecular Approach*. 2000, USA: Boston University.

16. Wolf, C. and P.J. Quinn, *Membrane lipid homeostasis.* Subcell Biochem, 2004. **37**: p. 317-57.

17. Herker, E. and M. Ott, *Emerging role of lipid droplets in host/pathogen interactions.* J Biol Chem, 2012. **287**(4): p. 2280-7.

18. Jarlier, V. and H. Nikaido, *Mycobacterial cell wall: structure and role in natural resistance to antibiotics.* FEMS Microbiol Lett, 1994. **123**(1-2): p. 11-8.

19. Daffe, M. and P. Draper, *The envelope layers of mycobacteria with reference to their pathogenicity.* Adv Microb Physiol, 1998. **39**: p. 131-203.

20. Raetz, C.R. and C. Whitfield, *Lipopolysaccharide endotoxins.* Annu Rev Biochem, 2002. **71**: p. 635-700.

21. Houben, E.N., L. Nguyen, and J. Pieters, *Interaction of pathogenic mycobacteria with the host immune system.* Curr Opin Microbiol, 2006. **9**(1): p. 76-85.

22. Bloom, B.R., *Tuberculosis: pathogenesis, protection, and control.* 1994.

23. WHO, *Tuberculosis report.* 2012.

24. Fauci, A.S., *Host factors and the pathogenesis of HIV-induced disease.* Nature, 1996. **384**(6609): p. 529-34.

25. Brennan, P.J. and H. Nikaido, *The envelope of mycobacteria.* Annu Rev Biochem, 1995. **64**: p. 29-63.

26. Ben Sidders, N.G.S., *Mycobacteria: Biology*. 2007, US: Wiley.

27. Cook, G.M., et al., *Physiology of mycobacteria.* Adv Microb Physiol, 2009. **55**: p. 81-182, 318-9.

28. Honer zu Bentrup, K. and D.G. Russell, *Mycobacterial persistence: adaptation to a changing environment.* Trends Microbiol, 2001. **9**(12): p. 597-605.

29. Brennan, P.J. and D.C. Crick, *The cell-wall core of Mycobacterium tuberculosis in the context of drug discovery.* Curr Top Med Chem, 2007. **7**(5): p. 475-88.

30. Nalin Rastogi, C.F., Hugo L. David, *Triple-layered structure of mycobacterial cell wall: Evidence for the existence of a polysaccharide-rich outer layer in 18 mycobacterial species.* Current Microbiology, 1986. **13**(5): p. 237.

31. J, P., *Electron micrograph of Mycobacterium tuberculosis*.

32. Harding, C.V. and W.H. Boom, *Regulation of antigen presentation by Mycobacterium tuberculosis: a role for Toll-like receptors.* Nat Rev Microbiol, 2010. **8**(4): p. 296-307.

33. Minnikin, D.E., et al., *The methyl-branched fortifications of Mycobacterium tuberculosis.* Chem Biol, 2002. **9**(5): p. 545-53.

34. Taneja, R., U. Malik, and G.K. Khuller, *Effect of growth temperature on the lipid composition of Mycobacterium smegmatis ATCC 607.* J Gen Microbiol, 1979. **113**(2): p. 413-6.

35. Gilleron, M., B. Lindner, and G. Puzo, *MS/MS approach for characterization of the fatty acid distribution on mycobacterial phosphatidyl-myo-inositol mannosides.* Anal Chem, 2006. **78**(24): p. 8543-8.

36. Amate, L., M. Ramirez, and A. Gil, *Positional analysis of triglycerides and phospholipids rich in long-chain polyunsaturated fatty acids.* Lipids, 1999. **34**(8): p. 865-71.

37. Nandedkar, A.K., *Fatty acid composition of mycobacterial lipids as an index of pathogenicity.* J Natl Med Assoc, 1982. **74**(12): p. 1191-3.

38. Khuller, G.K., et al., *Lipid composition and virulence of Mycobacterium tuberculosis H37Rv.* Aust J Exp Biol Med Sci, 1982. **60 (Pt 5)**: p. 541-7.

39. Goren, M.B., *Mycobacterial lipids: selected topics.* Bacteriol Rev, 1972. **36**(1): p. 33-64.

40. Morita, Y.S., et al., *Stress-induced synthesis of phosphatidylinositol 3-phosphate in mycobacteria.* J Biol Chem, 2010. **285**(22): p. 16643-50.

41. Sturm, A., et al., *Manipulation of host hepatocytes by the malaria parasite for delivery into liver sinusoids.* Science, 2006. **313**(5791): p. 1287-90.

42. Vissa, V.D. and P.J. Brennan, *The genome of Mycobacterium leprae: a minimal mycobacterial gene set.* Genome Biol, 2001. **2**(8): p. REVIEWS1023.

43. Brennan, P.J., *Structure, function, and biogenesis of the cell wall of Mycobacterium tuberculosis.* Tuberculosis (Edinb), 2003. **83**(1-3): p. 91-7.

44. Jankute, M., et al., *Arabinogalactan and lipoarabinomannan biosynthesis: structure, biogenesis and their potential as drug targets.* Future Microbiol, 2012. **7**(1): p. 129-47.

45. Kaur, D., et al., *New insights into the biosynthesis of mycobacterial lipomannan arising from deletion of a conserved gene.* J Biol Chem, 2007. **282**(37): p. 27133-40.

46. Khoo, K.H., et al., *Structural definition of acylated phosphatidylinositol mannosides from Mycobacterium tuberculosis: definition of a common anchor for lipomannan and lipoarabinomannan.* Glycobiology, 1995. **5**(1): p. 117-27.

47. Guerardel, Y., et al., *Structural study of lipomannan and lipoarabinomannan from Mycobacterium chelonae. Presence of unusual components with alpha 1,3-mannopyranose side chains.* J Biol Chem, 2002. **277**(34): p. 30635-48.

48. Khoo, K.H., et al., *Inositol phosphate capping of the nonreducing termini of lipoarabinomannan from rapidly growing strains of Mycobacterium.* J Biol Chem, 1995. **270**(21): p. 12380-9.

49. Chatterjee, D., et al., *Structural definition of the non-reducing termini of mannose-capped LAM from Mycobacterium tuberculosis through selective enzymatic degradation and fast atom bombardment-mass spectrometry.* Glycobiology, 1993. **3**(5): p. 497-506.

50. Welin, A., et al., *Incorporation of Mycobacterium tuberculosis lipoarabinomannan into macrophage membrane rafts is a prerequisite for the phagosomal maturation block.* Infect Immun, 2008. **76**(7): p. 2882-7.

51. Vignal, C., et al., *Lipomannans, but not lipoarabinomannans, purified from Mycobacterium chelonae and Mycobacterium kansasii induce TNF-alpha and IL-8 secretion by a CD14-toll-like receptor 2-dependent mechanism.* J Immunol, 2003. **171**(4): p. 2014-23.

52. Cole, S.T., et al., *Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.* Nature, 1998. **393**(6685): p. 537-44.

53. Gago, G., et al., *Fatty acid biosynthesis in actinomycetes.* FEMS Microbiol Rev, 2011. **35**(3): p. 475-97.

54. Gornicki, P., *Apicoplast fatty acid biosynthesis as a target for medical intervention in apicomplexan parasites.* Int J Parasitol, 2003. **33**(9): p. 885-96.

55. Goodman, C.D. and G.I. McFadden, *Fatty acid biosynthesis as a drug target in apicomplexan parasites.* Curr Drug Targets, 2007. **8**(1): p. 15-30.

56. Coombs, G.H. and S. Muller, *Recent advances in the search for new anti-coccidial drugs.* Int J Parasitol, 2002. **32**(5): p. 497-508.

57. Bachhawat, N. and S.C. Mande, *Identification of the INO1 gene of Mycobacterium tuberculosis H37Rv reveals a novel class of inositol-1-phosphate synthase enzyme.* J Mol Biol, 1999. **291**(3): p. 531-6.

58. Movahedzadeh, F., et al., *Inositol monophosphate phosphatase genes of Mycobacterium tuberculosis.* BMC Microbiol, 2010. **10**: p. 50.

59. Jackson, M., D.C. Crick, and P.J. Brennan, *Phosphatidylinositol is an essential phospholipid of mycobacteria.* J Biol Chem, 2000. **275**(39): p. 30092-9.

60. Guerin, M.E., et al., *Molecular recognition and interfacial catalysis by the essential phosphatidylinositol mannosyltransferase PimA from mycobacteria.* J Biol Chem, 2007. **282**(28): p. 20705-14.

61. Kordulakova, J., et al., *Identification of the required acyltransferase step in the biosynthesis of the phosphatidylinositol mannosides of mycobacterium species.* J Biol Chem, 2003. **278**(38): p. 36285-95.

62. Lea-Smith, D.J., et al., *Analysis of a new mannosyltransferase required for the synthesis of phosphatidylinositol mannosides and lipoarbinomannan reveals two lipomannan pools in corynebacterineae.* J Biol Chem, 2008. **283**(11): p. 6773-82.

63. Kremer, L., et al., *Characterization of a putative alpha-mannosyltransferase involved in phosphatidylinositol trimannoside biosynthesis in Mycobacterium tuberculosis.* Biochem J, 2002. **363**(Pt 3): p. 437-47.

64. Liu, J. and A. Mushegian, *Three monophyletic superfamilies account for the majority of the known glycosyltransferases.* Protein Sci, 2003. **12**(7): p. 1418-31.

65. Mishra, A.K., et al., *Identification of an alpha(1-->6) mannopyranosyltransferase (MptA), involved in Corynebacterium glutamicum lipomanann biosynthesis, and identification of its orthologue in Mycobacterium tuberculosis.* Mol Microbiol, 2007. **65**(6): p. 1503-17.

66. Mishra, A.K., et al., *Lipoarabinomannan biosynthesis in Corynebacterineae: the interplay of two alpha(1-->2)-mannopyranosyltransferases MptC and MptD in mannan branching.* Mol Microbiol, 2011. **80**(5): p. 1241-59.

67. A.K. Mishra, S.B., L.J. Alderwick, K. Futterer, G.S. Besra, *Chapter 6 Bacterial Lipoarabinomannan: Structure to Biogenesis in book Bacterial Glycomics: Cuttent Research, Technology and Applications*. 2011, UK: Caister Academic Press.

68. Skovierova, H., et al., *AftD, a novel essential arabinofuranosyltransferase from mycobacteria.* Glycobiology, 2009. **19**(11): p. 1235-47.

69. Alderwick, L.J., et al., *Identification of a novel arabinofuranosyltransferase (AftA) involved in cell wall arabinan biosynthesis in Mycobacterium tuberculosis.* J Biol Chem, 2006. **281**(23): p. 15653-61.

70. Birch, H.L., et al., *Biosynthesis of mycobacterial arabinogalactan: identification of a novel alpha(1-->3) arabinofuranosyltransferase.* Mol Microbiol, 2008. **69**(5): p. 1191-206.

71. Sulzenbacher, G., et al., *LppX is a lipoprotein required for the translocation of phthiocerol dimycocerosates to the surface of Mycobacterium tuberculosis.* EMBO J, 2006. **25**(7): p. 1436-44.

72. Cao, B., J.M. White, and S.J. Williams, *Synthesis of glycoconjugate fragments of mycobacterial phosphatidylinositol mannosides and lipomannan.* Beilstein J Org Chem, 2011. **7**: p. 369-77.

73. Scholl, D., et al., *An engineered R-type pyocin is a highly specific and sensitive bactericidal agent for the food-borne pathogen Escherichia coli O157:H7.* Antimicrob Agents Chemother, 2009. **53**(7): p. 3074-80.

74. Neyrolles, O. and C. Guilhot, *Recent advances in deciphering the contribution of Mycobacterium tuberculosis lipids to pathogenesis.* Tuberculosis (Edinb), 2011. **91**(3): p. 187-95.

75. Medzhitov, R., *Toll-like receptors and innate immunity.* Nat Rev Immunol, 2001. **1**(2): p. 135-45.

76. Akira, S., S. Uematsu, and O. Takeuchi, *Pathogen recognition and innate immunity.* Cell, 2006. **124**(4): p. 783-801.

77. Buwitt-Beckmann, U., et al., *TLR1- and TLR6-independent recognition of bacterial lipopeptides.* J Biol Chem, 2006. **281**(14): p. 9049-57.

78. Barral, D.C. and M.B. Brenner, *CD1 antigen presentation: how it works.* Nat Rev Immunol, 2007. **7**(12): p. 929-41.

79. Camus, J.C., et al., *Re-annotation of the genome sequence of Mycobacterium tuberculosis H37Rv.* Microbiology, 2002. **148**(Pt 10): p. 2967-73.

80. Kovacs-Simon, A., R.W. Titball, and S.L. Michell, *Lipoproteins of bacterial pathogens.* Infect Immun, 2011. **79**(2): p. 548-61.

81. Hutchings, M.I., et al., *Lipoprotein biogenesis in Gram-positive bacteria: knowing when to hold 'em, knowing when to fold 'em.* Trends Microbiol, 2009. **17**(1): p. 13-21.

82. Sutcliffe, I.C. and D.J. Harrington, *Pattern searches for the identification of putative lipoprotein genes in Gram-positive bacterial genomes.* Microbiology, 2002. **148**(Pt 7): p. 2065-77.

83. Sutcliffe, I.C. and D.J. Harrington, *Lipoproteins of Mycobacterium tuberculosis: an abundant and functionally diverse class of cell envelope components.* FEMS Microbiol Rev, 2004. **28**(5): p. 645-59.

84. Seshadri, C., et al., *Lipoproteins are major targets of the polyclonal human T cell response to Mycobacterium tuberculosis.* J Immunol, 2013. **190**(1): p. 278-84.

85. Lancioni, C.L., et al., *Mycobacterium tuberculosis lipoproteins directly regulate human memory CD4(+) T cell activation via Toll-like receptors 1 and 2.* Infect Immun, 2011. **79**(2): p. 663-73.

86. Banaei, N., et al., *Lipoprotein processing is essential for resistance of Mycobacterium tuberculosis to malachite green.* Antimicrob Agents Chemother, 2009. **53**(9): p. 3799-802.

87. Rampini, S.K., et al., *LspA inactivation in Mycobacterium tuberculosis results in attenuation without affecting phagosome maturation arrest.* Microbiology, 2008. **154**(Pt 10): p. 2991-3001.

88. Tschumi, A., et al., *Functional analyses of mycobacterial lipoprotein diacylglyceryl transferase and comparative secretome analysis of a mycobacterial lgt mutant.* J Bacteriol, 2012. **194**(15): p. 3938-49.

89. Rezwan, M., et al., *Lipoprotein synthesis in mycobacteria.* Microbiology, 2007. **153**(Pt 3): p. 652-8.

90. Drage, M.G., et al., *TLR2 and its co-receptors determine responses of macrophages and dendritic cells to lipoproteins of Mycobacterium tuberculosis.* Cell Immunol, 2009. **258**(1): p. 29-37.

91. Rahman, M.J., et al., *Impact of toll-like receptor 2 deficiency on immune responses to mycobacterial antigens.* Infect Immun, 2011. **79**(11): p. 4649-56.

92. Jayawardena-Wolf, J. and A. Bendelac, *CD1 and lipid antigens: intracellular pathways for antigen presentation.* Curr Opin Immunol, 2001. **13**(1): p. 109-13.

93. Peter G. Gibson, M.A., Richard Wood-Baker, Jimmy Volmink, Michael Hensley, Ulrich Costabel, *Evidence-Based Respiratory Medicine*. 2008: Wiley.

94. Jones, B.W., et al., *Different Toll-like receptor agonists induce distinct macrophage responses.* J Leukoc Biol, 2001. **69**(6): p. 1036-44.

95. Sassetti, C.M. and E.J. Rubin, *Genetic requirements for mycobacterial survival during infection.* Proc Natl Acad Sci U S A, 2003. **100**(22): p. 12989-94.

96. Drage, M.G., et al., *Mycobacterium tuberculosis lipoprotein LprG (Rv1411c) binds triacylated glycolipid agonists of Toll-like receptor 2.* Nat Struct Mol Biol, 2010. **17**(9): p. 1088-95.

97. Pecora, N.D., et al., *Mycobacterium tuberculosis LprA is a lipoprotein agonist of TLR2 that regulates innate immunity and APC function.* J Immunol, 2006. **177**(1): p. 422-9.

98. Jin, M.S., et al., *Crystal structure of the TLR1-TLR2 heterodimer induced by binding of a tri-acylated lipopeptide.* Cell, 2007. **130**(6): p. 1071-82.

99. Schwede, T., et al., *SWISS-MODEL: An automated protein homology-modeling server.* Nucleic Acids Res, 2003. **31**(13): p. 3381-5.

100. Organization, W.H., *World Malaria Report 2012*, 2012.

101. Teklehaimanot, A. and P. Mejia, *Malaria and poverty.* Ann N Y Acad Sci, 2008. **1136**: p. 32-7.

102. CDC, *Malaria Facts.*

103. Goldberg, D.E., R.F. Siliciano, and W.R. Jacobs, Jr., *Outwitting evolution: fighting drug-resistant TB, malaria, and HIV*. Cell, 2012. **148**(6): p. 1271-83.

104. McFadden, G.I., et al., *Plastid in human parasites.* Nature, 1996. **381**(6582): p. 482.

105. Lim, L. and G.I. McFadden, *The evolution, metabolism and functions of the apicoplast.* Philos Trans R Soc Lond B Biol Sci, 2010. **365**(1541): p. 749-63.

106. Fichera, M.E. and D.S. Roos, *A plastid organelle as a drug target in apicomplexan parasites.* Nature, 1997. **390**(6658): p. 407-9.

107. McFadden, G.I. and D.S. Roos, *Apicomplexan plastids as drug targets.* Trends Microbiol, 1999. **7**(8): p. 328-33.

108. Waller, R.F., et al., *Protein trafficking to the plastid of Plasmodium falciparum is via the secretory pathway.* EMBO J, 2000. **19**(8): p. 1794-802.

109. Mazumdar, J. and B. Striepen, *Make it or take it: fatty acid metabolism of apicomplexan parasites.* Eukaryot Cell, 2007. **6**(10): p. 1727-35.

110. Waller, R.F., et al., *A type II pathway for fatty acid biosynthesis presents drug targets in Plasmodium falciparum.* Antimicrob Agents Chemother, 2003. **47**(1): p. 297-301.

111. McMurry, L.M., M. Oethinger, and S.B. Levy, *Triclosan targets lipid synthesis.* Nature, 1998. **394**(6693): p. 531-2.

112. Waller, R.F., et al., *Nuclear-encoded proteins target to the plastid in Toxoplasma gondii and Plasmodium falciparum.* Proc Natl Acad Sci U S A, 1998. **95**(21): p. 12352-7.

113. McLeod, R., et al., *Triclosan inhibits the growth of Plasmodium falciparum and Toxoplasma gondii by inhibition of apicomplexan Fab I.* Int J Parasitol, 2001. **31**(2): p. 109-13.

114. Surolia, N. and A. Surolia, *Triclosan offers protection against blood stages of malaria by inhibiting enoyl-ACP reductase of Plasmodium falciparum.* Nat Med, 2001. **7**(2): p. 167-73.

115. Henry, M. and L. Debarbieux, *Tools from viruses: bacteriophage successes and beyond.* Virology, 2012. **434**(2): p. 151-61.

116. Ackermann, H.W., *Tailed bacteriophages: the order caudovirales.* Adv Virus Res, 1998. **51**: p. 135-201.

117. Aksyuk, A.A. and M.G. Rossmann, *Bacteriophage assembly.* Viruses, 2011. **3**(3): p. 172-203.

118. Orlova, E.V., *How viruses infect bacteria?* EMBO J, 2009. **28**(7): p. 797-8.

119. Nancy Trun, J.T., *Fundamental Bacterial Genetics*. 2009.

120. Ackermann, H.W., *Phage classification and characterization.* Methods Mol Biol, 2009. **501**: p. 127-40.

121. Ackermann, H.W., *Bacteriophage observations and evolution.* Res Microbiol, 2003. **154**(4): p. 245-51.

122. Grange, J.M. and R.G. Bird, *The nature and incidence of lysogeny in Mycobacterium fortuitum.* J Med Microbiol, 1975. **8**(2): p. 215-23.

123. Abedon, S.T. and J. Yin, *Bacteriophage plaques: theory and analysis.* Methods Mol Biol, 2009. **501**: p. 161-74.

124. Yu, F. and S. Mizushima, *Roles of lipopolysaccharide and outer membrane protein OmpC of Escherichia coli K-12 in the receptor function for bacteriophage T4.* J Bacteriol, 1982. **151**(2): p. 718-22.

125. Leiman, P.G., et al., *Morphogenesis of the T4 tail and tail fibers.* Virol J, 2010. **7**: p. 355.

126. Browning, C., et al., *Phage pierces the host cell membrane with the iron-loaded spike.* Structure, 2012. **20**(2): p. 326-39.

127. Chen, J., et al., *Defects in glycopeptidolipid biosynthesis confer phage I3 resistance in Mycobacterium smegmatis.* Microbiology, 2009. **155**(Pt 12): p. 4050-7.

128. Pedulla, M.L., et al., *Origins of highly mosaic mycobacteriophage genomes.* Cell, 2003. **113**(2): p. 171-82.

129. Hatfull, G.F., *Mycobacteriophages: genes and genomes.* Annu Rev Microbiol, 2010. **64**: p. 331-56.

130. Bartual, S.G., et al., *Structure of the bacteriophage T4 long tail fiber receptor-binding tip.* Proc Natl Acad Sci U S A, 2010. **107**(47): p. 20287-92.

131. Lee, S., et al., *Bxz1, a new generalized transducing phage for mycobacteria.* FEMS Microbiol Lett, 2004. **241**(2): p. 271-6.

132. Hatfull, G.F., et al., *Comparative genomic analysis of 60 Mycobacteriophage genomes: genome clustering, gene acquisition, and gene size.* J Mol Biol, 2010. **397**(1): p. 119-43.

133. Garcia-Doval, C. and M.J. van Raaij, *Structure of the receptor-binding carboxy-terminal domain of bacteriophage T7 tail fibers.* Proc Natl Acad Sci U S A, 2012. **109**(24): p. 9390-5.

134. Nilsson, E.C., et al., *The GD1a glycan is a cellular receptor for adenoviruses causing epidemic keratoconjunctivitis.* Nat Med, 2011. **17**(1): p. 105-9.

APPENDIX A

THE CRYSTAL STRUCTURE OF LPRG

*Reprinted with permission from "*Mycobacterium tuberculosis lipoprotein LprG (Rv1411c) binds triacylated glycolipid agonists of Toll-like receptor 2* " by Drage, M. G., Tsai, H. C., Pecora, N. D., Moody, D. B., Boom, W. H., Sacchettini, J. C. and Harding, C. V. et al. 2010. Nature Structural and Molecular Biology, **17**(9): p. 1088-95., Copyright 2010 by Nature America, Inc.

nature structural & molecular biology

Mycobacterium tuberculosis lipoprotein LprG (*Rv1411c*) binds triacylated glycolipid agonists of Toll-like receptor 2

Michael G Drage[1,6], Han-Chun Tsai[2,6], Nicole D Pecora[1,6], Tan-Yun Cheng[3], Ahmad R Arida[1,4], Supriya Shukla[1], Roxana E Rojas[5], Chetan Seshadri[3], D Branch Moody[3], W Henry Boom[5,7], James C Sacchettini[2,7] & Clifford V Harding[1,7]

Knockout of *lprG* results in decreased virulence of *Mycobacterium tuberculosis* (MTB) in mice. MTB lipoprotein LprG has TLR2 agonist activity, which is thought to be dependent on its N-terminal triacylation. Unexpectedly, here we find that nonacylated LprG retains TLR2 activity. Moreover, we show LprG association with triacylated glycolipid TLR2 agonists lipoarabinomannan, lipomannan and phosphatidylinositol mannosides (which share core structures). Binding of triacylated species was specific to LprG (not LprA) and increased LprG TLR2 agonist activity; conversely, association of glycolipids with LprG enhanced their recognition by TLR2. The crystal structure of LprG in complex with phosphatidylinositol mannoside revealed a hydrophobic pocket that accommodates the three alkyl chains of the ligand. In conclusion, we demonstrate a glycolipid binding function of LprG that enhances recognition of triacylated MTB glycolipids by TLR2 and may affect glycolipid assembly or transport for bacterial cell wall biogenesis.

Mycobacterium tuberculosis (MTB) infects one third of the world's population and continues to be a leading cause of death. Although immune responses by immunocompetent individuals can contain infection, sterilizing immunity is not usually achieved[1]. Infected individuals harbor latent disease with the potential for future reactivation, clinical disease, infectious spread and mortality.

The hydrophobic cell envelope of MTB may be involved in several aspects of tuberculosis pathogenesis, including long-term survival in the host. MTB cell wall components stimulate host responses and contribute to the activity of Freund's adjuvant[2]. The mycobacterial cell wall contains glycolipids, which contribute to resistance to bactericidal free radicals[3] and modulate immune functions, including phagosome maturation[4,5] and cytokine production. The cell wall also contains an abundance of N-terminally triacylated lipoproteins[6,7]. Four small homologous lipoproteins (LprG, LprA, LppX and LprF), only found in the suborder of Corynebacterineae, contain a signal peptide for secretion through the Sec system and a lipobox motif for lipid modification on a conserved cysteine. Diacylglycerol is linked by a thioester bond to the cysteine, and a third acyl chain is attached by an amide bond to the amino group of the cysteine, resulting in triacylation. These lipoproteins are predicted to be localized to the periplasm or cell wall—for example, anchored to the outer leaflet of the cell membrane through their acyl chains. LprG (Rv1411c) may function with the other protein in its operon, a membrane pump (Rv1410c). Knockout of the LprG or its operon results in attenuated growth and survival in mice and macrophages[8–10]. Deletion of the LprG operon in *Mycobacterium smegmatis* results in decreased sliding motility and altered cell morphology[11], suggesting that LprG function may be related to cell wall biosynthesis. LppX has also been proposed to be involved in cell wall biosynthesis by binding and transporting phthiocerol dimycocerosate (PDIM)[12].

Toll-like receptor 2 (TLR2), which forms heterodimers with TLR1 or TLR6, is an important contributor to innate immune recognition of MTB[13–23]. TLR2/TLR1 heterodimers bind triacylated lipopeptides. TLR2 agonist activity has been shown for the following MTB lipoproteins (MTB H37Rv gene nomenclature and protein name synonyms in parentheses): LpqH (*Rv3763*, 19-kD lipoprotein)[18–21], LprA (*Rv1270c*)[23], LprG (*Rv1411c*, p27)[24] and Psts1 (*Rv0934*, PhoS1 or p38)[25]. Mycobacterial agonists of TLR2 also include glycolipids—for example, phosphatidyl-(myo)-inositol mannosides (PIMs), lipomannans (LMs), lipoarabinomannans (LAMs) and inositol phosphate-capped LAMs (PI-LAMs)[13,14,22].

We investigated the structural basis of TLR2 agonist activity of MTB LprG. Crystal structures show that the thioether-linked diacylglycerol binds a hydrophobic pocket in TLR2, and the amide-linked third acyl chain binds TLR1 (ref. 26). Unexpectedly, our studies show that nonacylated LprG (NA-LprG) retains TLR2 stimulatory capacity, and the crystal structure of NA-LprG reveals a glycolipid binding pocket lined with hydrophobic residues that could accommodate lipids with three acyl chains. This pocket noncovalently binds triacylated glycolipids, and the introduction of a single point mutation in this pocket blocks the glycolipid binding function of LprG. We propose

[1]Department of Pathology, Case Western Reserve University/University Hospitals Case Medical Center, Cleveland, Ohio, USA. [2]Department of Biochemistry and Biophysics, Texas A&M University, College Station, Texas, USA. [3]Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA. [4]Department of Biochemistry, Case Western Reserve University/University Hospitals Case Medical Center, Cleveland, Ohio, USA. [5]Division of Infectious Diseases and Tuberculosis Research Unit, Case Western Reserve University/University Hospitals Case Medical Center, Cleveland, Ohio, USA. [6]These authors contributed equally to this work. [7]These authors contributed equally to this work. Correspondence should be addressed to C.V.H. (cvh3@cwru.edu).

Received 24 January; accepted 11 June; published online 8 August 2010; doi:10.1038/nsmb.1869

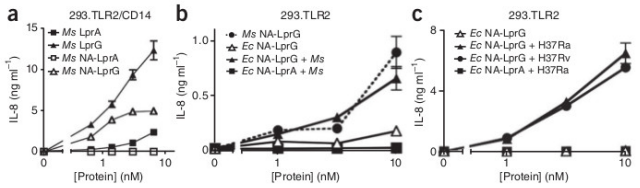


Figure 1 NA-LprG carries a mycobacterial TLR2 agonist. (**a**) HEK293.TLR2/CD14 cells show a dose-dependent IL-8 response to LprA, LprG and NA-LprG but no response to NA-LprA. Control HEK 293 cells lacking TLR2 and CD14 failed to respond to all four proteins (data not shown). The absence of CD14 (HEK293.TLR2 cells) reduced the apparent potency of NA-LprG but not acylated LprG or LprA (**Supplementary Fig. 1**). (**b**,**c**) NA-LprG can acquire TLR2 agonist activity from mycobacterial lysates. NA-LprG and NA-LprA were expressed in *E. coli* (*Ec*), purified by nickel-affinity and anion-exchange chromatography, incubated with control buffer or a lysate of *M. smegmatis* (*Ms*) (**b**), MTB H37Ra (**c**) or MTB H37Rv (**c**), repurified by nickel-affinity (**b**) or nickel-affinity and ion-exchange chromatography (**c**) and incubated with HEK293.TLR2 cells for 12 h. Minor technical differences between the panels resulted in different plateau IL-8 levels, but this was not due to intrinsic differences in activities of materials from *M. smegmatis* versus MTB (**Fig. 3** and data not shown). HEK293.TLR2/CD14 cells used in panel **a** give higher IL-8 secretion responses than HEK293.TLR2 cells used in **b** and **c** (see **Supplementary Fig. 1**). For all data panels, IL-8 production was quantified by ELISA, and data are reported as the mean ± s.d. of triplicate HEK293.TLR2 assays. Results are representative of at least three independent experiments.

that LprG functions in mycobacteria as a carrier of glycolipids during their trafficking and delivery to the mycobacterial cell wall, contributing to virulence[8,11] and providing potential opportunities for targeting in drug design. In addition, the glycolipid carrier function of LprG may facilitate recognition of triacylated glycolipids by TLR2.

RESULTS

LprG carries a mycobacterial TLR2 agonist

In the prolipoprotein maturation pathway, N-terminal acylation of a cysteine results in a lipoprotein capable of inducing a potent TLR2 response. To determine the importance of N-terminal acylation to TLR2 agonist activity, we compared the TLR2 agonist activity of two homologous MTB lipoproteins, LprG and LprA, with and without N-terminal triacylation. These lipoproteins are predicted to be similarly acylated by a common enzymatic pathway[27] and, therefore, to have similar TLR2 activity, yet acylated LprG induced TLR2-dependent interleukin 8 (IL-8) secretion with more than ten-fold greater potency than acylated LprA in a bioassay with HEK 293 cells transfected to express TLR2 and CD14 (**Fig. 1a**) or TLR2 alone (**Supplementary Fig. 1**). Accordingly, we designed nonacylated forms with the signal peptide removed and the N-terminal cysteine replaced with methionine to investigate whether structures other than the acyl chains could affect TLR2 agonist activity. We expressed recombinant hexahistidine-tagged acylated and nonacylated versions of MTB LprA and MTB LprG in *M. smegmatis* and assessed them for TLR2 agonist activity. Consistent with prior data[23], NA-LprA lacked TLR2 activity, indicating that acylation of LprA was essential for its TLR2 activity (**Fig. 1a**). In contrast, NA-LprG retained substantial TLR2 activity (**Fig. 1a**), showing that LprG possesses a previously unknown determinant of TLR2 agonist activity independent of its N-terminal acylation.

NA-LprG had substantially reduced activity when expressed in *Escherichia coli* instead of *M. smegmatis* (**Fig. 1b**), suggesting that LprG may carry TLR2 agonist(s) that are present in mycobacteria but not *E. coli*. Furthermore, TLR2 activity of NA-LprG purified from *E. coli* was substantially increased following incubation with a lysate of *M. smegmatis* (**Fig. 1b**), MTB H37Ra (**Fig. 1c**) or MTB H37Rv (**Fig. 1c**). In contrast, NA-LprA purified from *E. coli* did not acquire TLR2 activity from mycobacterial lysates (**Fig. 1b**,**c**). These results show that NA-LprG binds a mycobacterial TLR2 agonist and can deliver it for recognition by TLR2.

LprG has a hydrophobic binding pocket for TLR2 agonists

We produced crystals for a truncation form of NA-LprG lacking N-terminal residues 1–35 (that is, the signal peptide and the triacylated cysteine) and C-terminal residues 232–236 (secondary-structure prediction indicated random coils for these sequences). This NA-LprG construct was a monomer in solution as determined by gel filtration (isolated as a single peak of ~22 kDa). We found the crystals to be in the *C*2 space group. We solved the structure using multiwavelength anomalous dispersion (MAD) methods and refined it to a resolution of 2.0 Å with two molecules in the asymmetry unit (**Table 1**). Both molecules (A and B) have large hydrophobic cavities with distinct entry portals.

The overall LprG structure consists of a single domain in an α/β fold with a β-sheet composed of 10 antiparallel strands on one side and 6 α-helices on the opposite side. Between the β-sheet and α-helices is a large cavity (~1,500 Å^3) (**Fig. 2a**). The entrance to the cavity is approximately 9 Å × 20 Å and lies near β3, β4 and β5. The lower part of the molecule as positioned in **Figure 2a** is a twisted β-sheet (β6, β7, β8, β9 and β10) at the end of the cavity with a narrow binding channel (**Fig. 2a**,**b**). In the central portion of the cavity, the β-sheet's concave face is toward the α-helices, forming the large cavity. The cavity and the portal are lined primarily with the side chains of hydrophobic residues (**Fig. 2b**), which supports the hypothesis that LprG can bind lipids in the cavity.

Mutation of the pocket blocks association of TLR2 agonists

We designed a site-directed mutant of LprG in which we replaced Val91, located at the portal, with a bulkier tryptophan that was predicted to partially occlude the binding cavity (**Fig. 3**). We expressed the mutant NA-LprG V91W in *M. smegmatis* and crystallized it. The overall r.m.s. deviation between wild-type NA-LprG and mutant NA-LprG V91W is ~1.0 Å, and the Luzzati coordination mean error is ~0.2 Å. The crystal structure of NA-LprG V91W shows that Trp91 shifted the portal wall inward by 3.5 Å compared to the wild type (**Fig. 3** and **Table 1**). The region between Leu73 and Leu76 was also shifted into the cavity by 0.5 Å to form van der Waals interactions with Trp91, further narrowing the portal. The loop between β3 and β4 and the loop between β5 and helix α2 were also relocated by ~2 Å from the entrance in response to the presence of the indole ring. Due to movement of these loops and the tryptophan mutation, the mutation reduced the size of the cavity entrance from 9 Å × 20 Å to 8 Å × 13 Å. The resulting cavity volume of NA-LprG V91W (1,200 Å^3) was substantially smaller than that of NA-LprG (1,500 Å^3). Thus, as predicted, the V91W mutation reduced the dimensions of the pocket entrance dimensions and the volume of the cavity.

The narrowed cavity entrance and smaller cavity of NA-LprG V91W were predicted to provide steric hindrance to limit substrate binding. Consistent with this hypothesis, NA-LprG V91W had substantially reduced TLR2 agonist activity relative to NA-LprG (**Fig. 3**). Mutations at two other sites in the hydrophobic pocket (V194R and V217F) also decreased TLR2 agonist activity of NA-LprG (data not

Table 1 Data collection, phasing and refinement statistics for MAD (SeMet) structures

	Native NA-LprG	SeMet NA-LprG			LprG V91W	LprG-PIM
Data collection						
Space group	*C*2	*C*2			$P2_1$	*C*2
Cell dimensions						
a, *b*, *c* (Å)	95.3, 71.9, 61.3	95.5, 72.3, 62.3			39.7, 56.0, 96.9	95.3, 71.9, 61.3
α, β, γ (°)	90.0, 106.3, 90.0	90.0, 106.8, 90.0			90.0, 99.5, 90.0	90.0, 106.3, 90.0
		Peak	Inflection	Remote		
Wavelength	1.5418	0.9796	0.9798	0.9537	0.9794	0.9795
Resolution (Å)	2.0 (2.07–2.00)	1.9 (1.97–1.90)	1.9 (1.97–1.90)	1.9 (1.97–1.90)	1.8 (1.82–1.8)	1.85 (1.92–1.85)
R_{sym}	0.068 (0.73)	0.076 (0.57)	0.057 (0.41)	0.061 (0.53)	0.063 (0.59)	0.068 (0.72)
$I / \sigma I$	461.1 (13.7)	42.5 (2.4)	80.3 (4.1)	55.0 (3.0)	42.7 (2.3)	256.6 (5.5)
Completeness (%)	98.7	90.6	98.1	94.5	97.0 (98.5)	99.9 (99.8)
Redundancy	7.5	6.6	7.2	6.9	2.8 (2.7)	5.7 (5.2)
Refinement						
Resolution (Å)	50.0–2.0				50.0–1.8	50.0–1.85
No. reflections	811,524				696,007	469,056
R_{work} / R_{free}	22.90 / 27.8				21.5 / 25.3	22.2 / 25.8
No. atoms						
Protein	2,874				3,030	2,876
Ligand / ion	–				–	97
Water	114				479	175
B-factors						
Protein	49.8				26.0	43.2
Ligand	–				–	63.2
Water	49.1				33.5	44.0
R.m.s. deviations						
Bond lengths (Å)	0.008				0.006	0.007
Bond angles (°)	1.152				1.058	1.144

Values in parentheses are for highest-resolution shell.

shown). Furthermore, when we expressed NA-LprG V91W in *E. coli*, purified it and then incubated it with a lysate of *M. smegmatis* or MTB, the V91W mutant lacked the ability to acquire TLR2 agonist activity from mycobacteria (**Fig. 3**). These results indicate that the hydrophobic cavity is a binding site for TLR2 agonist(s), which are delivered by LprG for recognition by TLR2.

Mycobacterial glycolipids are associated with LprG

We found NA-LprG to signal through TLR2-TLR1 heterodimers (**Supplementary Fig. 2**), a pattern observed with mycobacterial glycolipids such as LAM[16] and LM[15], as well as triacylated lipopeptide. To identify its associated TLR2 agonists, we purified NA-LprG from *M. smegmatis* by nickel-affinity and anion-exchange chromatography and subjected it to PAGE (**Fig. 4**). Silver stain and anti-hexahistidine western blotting of NA-LprG, NA-LprG V91W or NA-LprA shows a single band at approximately 24 kDa (**Fig. 4a** and data not shown). Staining of carbohydrates after periodate oxidation showed additional bands that comigrated with protein molecular weight markers of 25–35 kDa, 14–18 kDa and <10 kDa (**Fig. 4b**), which correspond roughly to the expected positions of LAM, LM and PIM, respectively, and were found to comigrate with authentic glycolipid standards (**Fig. 4d** and data not shown). These bands likely represent glycolipids because they were seen after periodate oxidation but not on conventional silver stain, and glycolipids in the PIM-LAM series resolve as broad bands based on heterogeneity of the arabinan and mannan components in each molecular species. Notably, LAM and LM were associated with NA-LprG but not NA-LprG V91W or LprA, indicating that these glycolipid agonists of TLR2 are associated preferentially with NA-LprG. We detected an association of PIM with NA-LprG, to a lesser degree with NA-LprA and only minimally with NA-LprG V91W. Western blotting with polyclonal anti–*M. bovis* BCG antibody revealed bands consistent with LAM, LM and PIM in association with NA-LprG but not NA-LprG V91W or NA-LprA (**Fig. 4c**; anti-BCG staining may detect only a subset of PIM species). Furthermore, western blotting with a LAM-specific monoclonal antibody confirmed the association of LAM with NA-LprG and its absence from NA-LprG V91W and NA-LprA (**Fig. 4d**).

LAM, LM, and PIM share a common structural core, suggesting that they all associate with NA-LprG via this shared structural motif, which may also contribute to TLR2 agonist activity. To determine the structures of small molecules associated with NA-LprG and related proteins, we used methanol to denature proteins and allow detection of protein-associated small molecules with mass to charge (m/z) ratio up to 2,000 by nESI-MS. Methanol alone (**Fig. 4e**) or methanol elutes of an unrelated protein, Pab C (**Supplementary Fig. 3**), did not

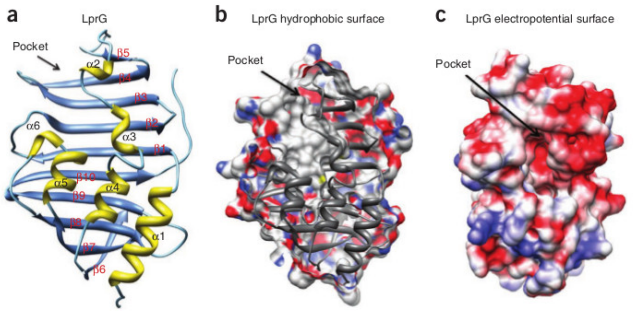


Figure 2 Crystal structure of NA-LprG reveals a hydrophobic pocket with the potential to carry a TLR2 agonist. (**a**) LprG structure viewed in ribbons. (**b**) LprG hydrophobic surface slab view clipped to cavity center (white, carbon; red, oxygen; blue, nitrogen; yellow, sulfur). (**c**) Electrostatic surface view of the LprG cavity entrance (red, negative; blue, positive; white, neutral). To show the entrance, LprG is rotated 90° to the right and 30° down relative to the other views. The cavity entrance is located between the β-sheet and α-helices.

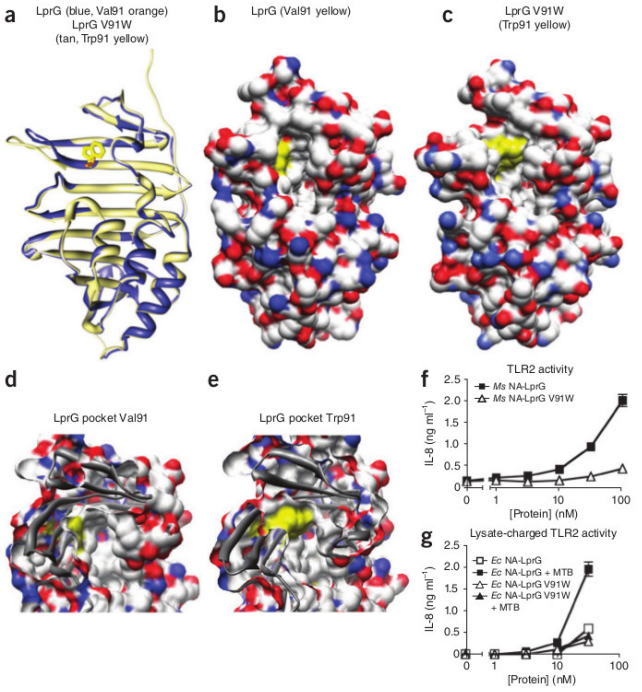


Figure 3 Site-directed mutagenesis to alter the structure of the NA-LprG pocket and binding of TLR2 agonists. (**a**) Superimposition ribbon view of NA-LprG (blue, Val91 in orange) and mutant NA-LprG V91W (tan, Trp91 in yellow). In addition to the V91W mutation, NA-LprG and NA-LprG V91W constructs used for crystallization included minor differences unrelated to the pocket structure (for example, NA-LprG V91W has a longer C-terminal coil). Otherwise, overall structures were similar with differences due to the V91W mutation localized to the cavity and entrance. The movement of β3, β4 and the loop between β3 and β4 affects cavity and entrance size. (**b**) Hydrophobic view of the NA-LprG entrance (white, carbon; red, oxygen; blue, nitrogen; yellow, Val91). (**c**) Hydrophobic view of NA-LprG V91W (yellow, Trp91). The extra C-terminal coil of NA-LprG V91W was removed for better surface comparison. The V91W mutation causes the cavity wall to shift by 3.5 Å. (**d**) View revealing the hydrophobic surface of the NA-LprG cavity. (**e**) NA-LprG V91W cavity. The V91W mutation narrows the pocket to reduce the space of ligand binding. (**f**) TLR2 activity of NA-LprG and NA-LprG V91W expressed in *M. smegmatis* (*Ms*) and tested on HEK293.TLR2 cells. (**g**) TLR2 activity of NA-LprG and NA-LprG V91W expressed in *E. coli* (*Ec*), purified, incubated with MTB H37Ra or H37Rv lysate and repurified. Data are reported as mean ± s.d. of triplicate HEK293.TLR2 assays.

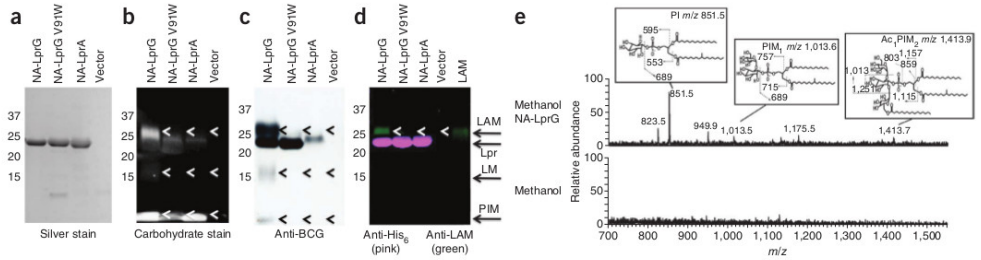
give detectable ions. In contrast, NA-LprG expressed in *M. smegmatis* yielded ions corresponding to mycobacterial phospholipids (**Fig. 4e** and **Supplementary Fig. 3**). Ions detected at *m/z* 851.4, 1,013.5, 1,175.5 and 1,413.7 corresponded to the expected masses of diacyl phosphatidylinositol, diacyl phosphatidylinositol monomannoside (PIM_1), diacyl phosphatidylinositol dimannoside (PIM_2) and triacyl PIM_2 (Ac_1PIM_2), respectively. CID-MS yielded product ions expected from these proposed structures, confirming assignments based on mass alone (**Fig. 4e** and **Supplementary Fig. 3**). For example, Ac_1PIM_2 (*m/z* 1,413.7) yielded products expected from the loss of mannose (*m/z* 1,251), loss of acyl mannose (*m/z* 1,013), loss of C16:0 acyl (*m/z* 1,157), loss of C19:0 fatty acyl (*m/z* 1,115) and acyl phosphoinositol dimannoside (*m/z* 803). Thus, LprG binds at least four structurally related molecules that all contain phosphatidylinositol as the core structure but differ in the number of mannose units and fatty acyl chains.

We used LC-MS to compare mass spectral signals of individual molecular species eluted from the same molar quantity of NA-LprG, NA-LprG V91W or NA-LprA (**Fig. 4f–i**). We used total ion current as a control to confirm comparable input of protein (**Fig. 4f**), and we monitored signals simultaneously in narrow ranges (±0.5 a.m.u.) near the expected masses of diacyl phosphatidylinositol (**Fig. 4g**), diacyl PIM_1 (**Fig. 4h**) and Ac_1PIM_2 (**Fig. 4i**). Diacyl phosphatidylinositol and diacyl PIM_1 were associated with NA-LprG and NA-LprA but only in lower amounts with NA-LprG V91W. In contrast, triacylated Ac_1PIM_2 was only associated with NA-LprG. Whereas LprA binds only diacylated PIMs, LprG binds both diacylated and triacylated PIMs. Because a large proportion of LAM is triacylated[28], the preferential ability of LprG to bind triacylated PIMs likely explains the greater association of LAM with LprG relative to LprA.

There are no known differences between *M. smegmatis* and MTB with regard to structures of phosphatidylinositol and PIM species. Comparative mass spectral profiling of phospholipids isolated from whole bacteria revealed the following m/z values for abundant molecular species in MTB versus *M. smegmatis*, respectively: 851.564 versus 851.562 for phosphatidylinositol, 1,013.66 versus 1,013.64 for PIM_2 and 1,413.89 versus 1,413.89 for Ac_1PIM_2. These data reveal no substantial differences in structures of phosphatidylinositol or PIM species between *M. smegmatis* and MTB, suggesting that these glycolipids will similarly associate with LprG in both *M. smegmatis* and MTB (LprG was cloned from MTB for these studies).

Structure of Ac_1PIM_2 bound to NA-LprG

To understand the structural basis of glycolipid binding, we incubated NA-LprG with an equimolar mixture of PIM and phosphatidylinositol from *M. smegmatis*. We solved the crystal structure of Ac_1PIM_2 bound to NA-LprG to 1.85-Å resolution (**Fig. 5** and **Table 1**). The overall structure of the protein, the electrostatic surface of the complex and the cavity volume were similar to those of the apo protein (r.m.s. deviation of 0.6 Å and cavity volume of 1,500 $Å^3$). In the NA-LprG PIM crystal, one molecule (A) in the asymmetric unit has a triacylated Ac_1PIM_2 bound, and the other molecule (B) has nothing bound. Although both protein molecules are very similar in structure (r.m.s. deviation of 1.3 Å), molecule B shows an ~2-Å movement for the α-helix–loop–α-helix motif at the portal to the Ac_1PIM_2 binding site. The movement of this α-helix–loop–α-helix motif (Asp94–Gln106) repositions the side chain of Tyr102 ~3 Å, where it sterically prohibits



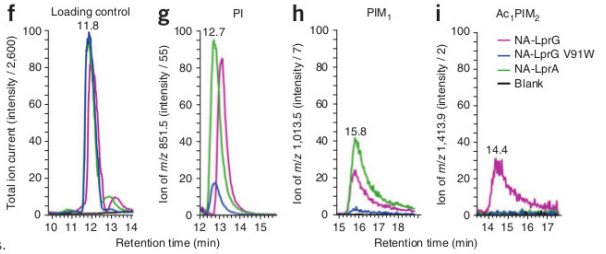


Figure 4 Triacylated mycobacterial glycolipids are associated with NA-LprG. (**a**–**d**) SDS-PAGE analysis of proteins purified from *M. smegmatis*. Samples were visualized by silver stain (**a**), Pro-Q stain for carbohydrates (**b**), polyclonal anti-BCG western blot (**c**) and monoclonal anti-hexahistidine (anti-His_6, pink) and anti-LAM (green) (**d**). The 10-kDa molecular weight marker ran at the gel dye front. Arrows on right labeled LAM, LM and PIM represent the positions observed with glycolipid standards. Arrowheads within the gels indicate the expected positions for LAM, LM and PIM. See **Supplementary Methods**. (**e**) Negative-mode ESI-MS of methanol-treated NA-LprG yielded ions corresponding in mass to mycobacterial phospholipids. Ions of *m/z* 851.5, 1,013.5 and 1,413.7 were subjected to negative-mode CID-MS, yielding ions that corresponded to masses and fragments indicated in the insets and **Supplementary Figure 3**. Ions detected near *m/z* 823.5 correspond those to an alternately acylated form of phosphatidylinositol, those near *m/z* 949.9 correspond to an H_3PO_4 adduction of phosphatidylinositol, and those near 1,175.5 correspond to diacylated PIM2. (**f**–**i**) LC-MS analysis was carried out on 4.3 nmol of NA-LprG (pink), NA-LprG V91W (blue), NA-LprA (green) and solvent blank (black). The total ion current trace shows that signals from nonlipidic components of protein preparations were detected at similar levels, serving as a loading control (**f**). Mass chromatograms measured in narrow mass ranges corresponding to the masses of phosphatidylinositol (*m/z* 851.5) (**g**) diacyl PIM_1 (*m/z* 1,013.6) (**h**) and triacyl Ac_1PIM_2 (*m/z* 1,413.76) (**i**) are shown.

Ac_1PIM_2 binding. The movements are not influenced by crystal packing, as the apo protein has nearly identical crystal packing.

Triacylated Ac_1PIM_2 (C19:0, C16:0, C16:0) was resolved in the binding cavity of molecule A (**Fig. 5a**,**c**). The interactions between LprG and Ac_1PIM_2 were mainly through van der Waals contacts between the hydrophobic side chains within the cavity and the three acyl chains of the ligand. The deepest part of the cavity, where the mannosyl C16 chain is located, appears to be capable of accommodating an

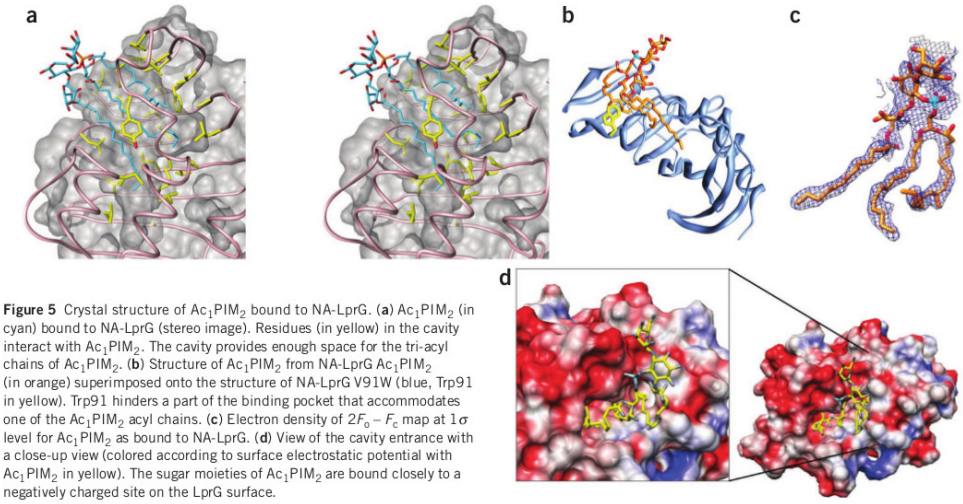


Figure 5 Crystal structure of Ac_1PIM_2 bound to NA-LprG. (**a**) Ac_1PIM_2 (in cyan) bound to NA-LprG (stereo image). Residues (in yellow) in the cavity interact with Ac_1PIM_2. The cavity provides enough space for the tri-acyl chains of Ac_1PIM_2. (**b**) Structure of Ac_1PIM_2 from NA-LprG Ac_1PIM_2 (in orange) superimposed onto the structure of NA-LprG V91W (blue, Trp91 in yellow). Trp91 hinders a part of the binding pocket that accommodates one of the Ac_1PIM_2 acyl chains. (**c**) Electron density of $2F_o - F_c$ map at 1σ level for Ac_1PIM_2 as bound to NA-LprG. (**d**) View of the cavity entrance with a close-up view (colored according to surface electrostatic potential with Ac_1PIM_2 in yellow). The sugar moieties of Ac_1PIM_2 are bound closely to a negatively charged site on the LprG surface.

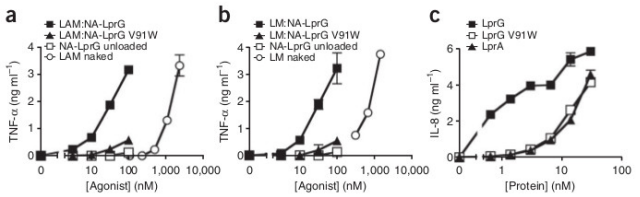


Figure 6 LprG binds purified mycobacterial glycolipids and facilitates their recognition by TLR2. (**a**,**b**) NA-LprG was purified from *E. coli*, incubated with LAM or LM from *M. smegmatis*, repurified and assessed for TLR2 activity using bone marrow–derived macrophages. TNF-α production was quantified by ELISA. LAM:NA-LprG, NA-LprG that was incubated with LAM and repurified; NA-LprG unloaded, NA-LprG that was sham-loaded with buffer and repurified; Naked LAM, glycolipid in the absence of NA-LprG. Data are reported as mean ± s.d. of triplicate macrophage assays. (**c**) Acylated LprG, LprA and LprG V91W were purified from *M. smegmatis* and assessed for TLR2 agonist activity with HEK293.TLR2/CD14 as in **Figure 1**. Data are reported as mean ± s.d. of triplicate HEK293.TLR2/CD14 assays.

acyl chain that is at least two more carbons in length (**Fig. 5a**). We found the sugar moieties of Ac_1PIM_2 on the edge of the entry portal, directly outside the cavity. The α-helix–loop–α-helix region occupies about half of the LprG cavity entrance, very close to where mannosyl inositol binds in molecule A. This region is negatively charged (**Fig. 5d**), providing a favorable environment for sugar binding. However, in the structure of NA-LprG PIM, we observed only two relatively weak interactions between the sugar moieties of Ac_1PIM_2. They are hydrogen bonds between the backbone oxygen of Gly42 and the 4-hydroxy of C16 acyl mannose (~2.7 Å) and between the backbone oxygen of Asp100 and the phosphate oxygen of the ligand (~3.0 Å). Therefore, the major component of ligand recognition involves the hydrophobic residues in the cavity.

Superimposed structures of the NA-LprG V91W mutant and NA-LprG PIM reveal that Trp91 overlaps with the C19 acyl chain at the sn1 position of Ac_1PIM_2 (**Fig. 5b**). Thus, the V91W mutation would occlude binding of Ac_1PIM_2, correlating with decreased association of Ac_1PIM_2 from MS (**Figs. 4** and **5**) and decreased TLR2 activity (**Fig. 3**).

LprG binds mycobacterial glycolipids that are TLR2 agonists

To directly test the ability of NA-LprG to bind mycobacterial glycolipids, we purified NA-LprG from *E. coli* (with little or no TLR2 activity), incubated it with LM or LAM, repurified it by nickel-affinity and anion-exchange chromatography and tested it for TLR2 agonist activity. NA-LprG was able to bind LM and LAM and deliver them for recognition by TLR2 (**Fig. 6a**,**b**), but NA-LprG V91W (**Fig. 6a**,**b**) and NA-LprA (data not shown) lacked this ability, consistent with the pattern of LM and LAM association in **Figure 4**. Notably, association with NA-LprG enhanced the apparent potency of LAM and LM by at least 1.5 log orders of magnitude (**Fig. 6a**,**b**). This comparison is based on the molar concentrations of NA-LprG shown in **Figure 6**, assuming that 100% of the NA-LprG molecules were loaded with LM or LAM. In the likelihood that glycolipid loading was less than 100% efficient, the factor by which association with NA-LprG enhanced glycolipid potency would be higher. We conclude that NA-LprG binds mycobacterial glycolipids and substantially enhances their recognition by TLR2.

Glycolipids contribute to the TLR2 activity of acylated LprG

Our studies indicate existence of two distinct determinants of TLR2 agonist activity of LprG: N-terminal acylation and chaperoned glycolipids. We studied acylated LprG and acylated LprG V91W to determine the relative contribution of these mechanisms. Acylated LprG V91W was more than a factor of 10 less potent than acylated LprG (**Fig. 6c**), indicating that the majority of TLR2 agonist activity of acylated LprG was attributable to carried glycolipid. Consistent with this finding, Western blot analysis revealed association of LAM with acylated LprG (**Supplementary Fig. 4**). Acylated LprG V91W had potency similar to that of acylated LprA (**Fig. 6c**), indicating that TLR2 activity remaining when glycolipids were not associated with LprG was similar to that of LprA, reflecting similar N-terminal acylation of these molecules. Thus, the higher potency of LprG relative to LprA results from LprG-associated glycolipids, which are a major determinant of the overall LprG TLR2 activity.

DISCUSSION

The discovery of a glycolipid carrier function has significant implications for the function of LprG in bacterial physiology. LprG is widely present in mycobacteria with 100% sequence identity among MTB complex species. LprG (*Rv1411c*) and p55 (*Rv1410c*) are encoded in an operon that is important for virulence[8,11,29]. We propose that LprG serves as a carrier to facilitate the assembly or trafficking of glycolipids. The LprG pocket structure confers specificity for binding triacylated lipids, which are a relatively small pool within the much larger pool of diacylated lipids. Glycolipids have important roles in bacterial physiology and host-pathogen interactions, and a potential future goal is the inhibition of these processes by targeting the pocket of LprG with chemotherapeutics.

LprG has a pocket that accommodates three acyl chains, raising the possibility that LprG, like LolA and LolB of *Escherichia* coli, may bind lipopeptides via their triacylated N termini. LolA and LolB, however, show less than 20% identity to LprG, and their lipid binding cavities are considerably smaller than LprG's. Furthermore, MS did not reveal other lipopeptides associated with LprG purified from *M. smegmatis*, possibly due to competition by glycolipid ligands in mycobacteria or potential structural constraints. It is also unlikely that LprG would bind its own acyl peptide without considerable unfolding as the N-terminal side of LprG is ~50 Å away from the binding cavity entrance. These considerations fit with functional data (**Fig. 6c**) and biochemical data (**Supplementary Fig. 4**) that show glycolipid association with both NA-LprG and acylated LprG.

NA-LprG is unique among the studied proteins in its ability to bind Ac_1PIM_2, LM and LAM (**Fig. 4**), all of which may contain the triacylated PIM core structure that fits within the LprG pocket (**Fig. 5**). LAM is a complex polymer with a high molecular weight that cannot be analyzed by the MS systems reported here, but other studies show that a large proportion of naturally occurring LAM is triacylated[28]. Therefore, the preferential association of LAM with NA-LprG as compared to NA-LprA or NA-LprG V91W may be explained by acyl chain interactions with the pocket, as seen in the NA-LprG Ac_1PIM_2 crystal structure. These findings are consistent with the observed TLR2/TLR1 dependence of signaling by MTB glycolipids in the PIM-LM-LAM series[14–17] and the concept that TLR2 activity of these mycobacterial glycolipids requires their triacylation[30].

Recognition of PIM by LprG primarily involves interactions between the hydrophobic cavity of LprG and the three fatty acyl chains of the glycolipid (**Fig. 5a**,**d**). We cannot exclude contributions of the longer carbohydrate chains of LM and LAM to LprG binding. Our MS data indicate that the structures of PIMs associated with LprG—for example, Ac_1PIM_2—are similar in MTB and *M. smegmatis*. LMs and LAMs are formed by additional glycosylation of PIMs. LAM is mannose capped in MTB, unlike *M. smegmatis*, and the resulting mannose-capped LAM has lower TLR2 potency than PI-LAM from *M. smegmatis*, but MTB LAM has substantial TLR2 agonist activity. Furthermore, the ability of MTB lysate to efficiently 'charge' LprG with TLR2 activity (**Fig. 1c**) indicates that MTB glycolipids are an effective source of LprG-chaperoned TLR2 agonist.

The binding pocket of TLR2 is hydrophobic with a volume of 1,800 Å3, close to that of the LprG binding pocket, but contains only two acyl chains of lipopeptides[26,31] and glycolipids[31], perhaps due to a restricted entrance. The third acyl chain of triacylated glycolipids may bind TLR1 (ref. 26). Thus, the triacyl structure that provides specificity for binding to LprG also determines the ability of MTB glycolipids to signal through TLR2/TLR1. Although LprA does not chaperone triacylated TLR2 agonists, modeling of LprA also reveals a hydrophobic pocket (**Supplementary Fig. 5**), and we observed an association of diacylated glycolipids with LprA. Another MTB lipoprotein, LppX, has a hydrophobic pocket[12]. Other mycobacterial lipoproteins may be carriers of hydrophobic ligands, some of which may be TLR2 agonists, and the potential contribution of chaperoned molecules to TLR2 activity of other lipoproteins should be considered.

Although the localizations of LprG and NA-LprG remain to be determined, acylated LprG is predicted to traffic to the outer surface of the mycobacterial cell membrane, where LM and LAM are assembled[32]. LprG may contribute to LAM assembly, trafficking and/or insertion into the cell wall. NA-LprG is predicted to be localized to the cytoplasm, as it lacks a signal peptide, and its loading with glycolipids may occur upon bacterial lysis. NA-LprG is a very useful tool to dissect the different components of LprG TLR2 activity, carried glycolipids and N-terminal acylation, but future studies of physiologic functions should focus on acylated LprG.

LprG may enhance TLR2 recognition of MTB by direct and/or indirect mechanisms. LprG may directly convey glycolipids to host receptors, for example, TLR2 or CD14. This mechanism implies release of LprG from MTB, either spontaneously or during phagosomal processing of MTB, which may allow LprG to reach TLR2 molecules that target to phagosomes. In addition, a role for LprG in the assembly, trafficking and cell-wall insertion of glycolipids indicates that LprG may affect the bioavailability of glycolipid TLR2 agonists.

Our results suggest that host cells can co-opt the carrier function of LprG to enhance TLR2 recognition of mycobacterial glycolipids. Association of glycolipids with NA-LprG increased their apparent potency by at least a factor of 1.5 log. To our knowledge, this is the first report of a pathogen-derived protein serving as a carrier for the delivery of hydrophobic agonists for innate immune recognition. For LPS, another hydrophobic TLR ligand, transfer to TLR4 is enhanced by host-expressed carriers, LPS-binding protein[33] and CD14 (ref. 34). LprG may deliver glycolipids to TLR2 via delivery to CD14, which we observed to increase recognition of NA-LprG, presumably by enhancing glycolipid delivery to TLR2.

LprG influences progression of MTB infection *in vivo*, as genetic deletion of *Rv1411c* (encoding LprG) in MTB H37Rv reduces the virulence of MTB in C57BL/6 mouse pulmonary aerosol infection models (CFU burden decreases by a factor of 1,000 to 10,000 relative to wild-type MTB H37Rv or a strain in which the genetic deletion has been complemented; Banaei, N., personal communication). This establishes the importance of LprG to virulence, apparently by its contributions to MTB cell-wall assembly, delivery of glycolipids to TLR2 or both.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/nsmb/.

Accession codes. Protein Data Bank: Coordinates and structure factors have been deposited under accession codes 3MHA (LprG-PIM), 3MH9 (LprG V91W) and 3MH8 (LprG).

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

We thank V. Anderson and L. Sweet for advice and assistance with MS, A.G. Hise (Case Western Reserve Univ.) and A. Shizuo (Osaka Univ.) for bone marrow from $TLR1^{-/-}$ $TLR6^{-/-}$ mice, X. Ding, N. Nagy and K. Daniel for technical assistance, T. Musa for comments on the manuscript and K. Dobos-Elder and J. Belisle (Colorado State Univ.) for MTB lysates, antibodies and the pVV16 vector. This work was supported by US National Institutes of Health (NIH) grants AI035726, AI034343 and AI069085 to C.V.H., HL055967 and AI027243 to W.H.B., AI071155 and AI049313 to D.B.M. and AI068135 to J.C.S., the Robert A. Welch Foundation (J.C.S.), the Irvington Institute Fellowship Program of the Cancer Research Institute (C.S.), American Lung Association grant RG48786N (R.E.R.) and the Burroughs Wellcome Fund for Translational Research (D.B.M. and C.S.). Core facilities of the Case Western Reserve University Center for AIDS Research were supported by NIH grant AI067093.

AUTHOR CONTRIBUTIONS

M.G.D., H.-C.T., N.D.P., T.-Y.C., A.R.A., S.S., R.E.R. and C.S. designed, performed and interpreted experiments and prepared the manuscript; D.B.M., W.H.B., J.C.S. and C.V.H. designed and interpreted experiments and prepared the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/nsmb/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Flynn, J.L. & Chan, J. Immunology of tuberculosis. *Annu. Rev. Immunol.* **19**, 93–129 (2001).
2. Janeway, C.A. Jr. Approaching the asymptote? Evolution and revolution in immunology. *Cold Spring Harb. Symp. Quant. Biol.* **54**, 1–13 (1989).
3. Chan, J. *et al.* Microbial glycolipids: possible virulence factors that scavenge oxygen radicals. *Proc. Natl. Acad. Sci. USA* **86**, 2453–2457 (1989).
4. Fratti, R.A., Chua, J., Vergne, I. & Deretic, V. *Mycobacterium tuberculosis* glycosylated phosphatidylinositol causes phagosome maturation arrest. *Proc. Natl. Acad. Sci. USA* **100**, 5437–5442 (2003).
5. Vergne, I. *et al. Mycobacterium tuberculosis* phagosome maturation arrest: mycobacterial phosphatidylinositol analog phosphatidylinositol mannoside stimulates early endosomal fusion. *Mol. Biol. Cell* **15**, 751–760 (2004).
6. Cole, S.T. *et al.* Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537–544 (1998).
7. Sutcliffe, I.C. & Harrington, D.J. Lipoproteins of *Mycobacterium tuberculosis*: an abundant and functionally diverse class of cell envelope components. *FEMS Microbiol. Rev.* **28**, 645–659 (2004).
8. Bigi, F. *et al.* The knockout of the lprG-Rv1410 operon produces strong attenuation of *Mycobacterium tuberculosis*. *Microbes Infect.* **6**, 182–187 (2004).
9. Rengarajan, J., Bloom, B.R. & Rubin, E.J. Genome-wide requirements for *Mycobacterium tuberculosis* adaptation and survival in macrophages. *Proc. Natl. Acad. Sci. USA* **102**, 8327–8332 (2005).
10. Sassetti, C.M. & Rubin, E.J. Genetic requirements for mycobacterial survival during infection. *Proc. Natl. Acad. Sci. USA* **100**, 12989–12994 (2003).
11. Farrow, M.F. & Rubin, E.J. Function of a mycobacterial major facilitator superfamily pump requires a membrane-associated lipoprotein. *J. Bacteriol.* **190**, 1783–1791 (2008).
12. Sulzenbacher, G. *et al.* LppX is a lipoprotein required for the translocation of phthiocerol dimycocerosates to the surface of *Mycobacterium tuberculosis*. *EMBO J.* **25**, 1436–1444 (2006).

13. Jones, B.W. *et al.* Different Toll-like receptor agonists induce distinct macrophage responses. *J. Leukoc. Biol.* **69**, 1036–1044 (2001).
14. Gilleron, M., Quesniaux, V.F. & Puzo, G. Acylation state of the phosphatidyl inositol hexamannosides from *Mycobacterium bovis* BCG and *Mycobacterium tuberculosis* H37Rv and its implication in TLR response. *J. Biol. Chem.* **278**, 29880–29889 (2003).
15. Elass, E. *et al.* Mycobacterial lipomannan induces matrix metalloproteinase-9 expression in human macrophagic cells through a toll-like receptor 1 (TLR1)/TLR2- and CD14-dependent mechanism. *Infect. Immun.* **73**, 7064–7068 (2005).
16. Tapping, R.I. & Tobias, P.S. Mycobacterial lipoarabinomannan mediates physical interactions between TLR1 and TLR2 to induce signaling. *J. Endotoxin Res.* **9**, 264–268 (2003).
17. Bhatt, K. & Salgame, P. Host innate immune response to *Mycobacterium tuberculosis*. *J. Clin. Immunol.* **27**, 347–362 (2007).
18. Pai, R.K., Convery, M., Hamilton, T.A., Boom, W.H. & Harding, C.V. Inhibition of IFN-γ-induced class II transactivator expression by a 19-kDa lipoprotein from *Mycobacterium tuberculosis*: a potential mechanism for immune evasion. *J. Immunol.* **171**, 175–184 (2003).
19. Noss, E.H. *et al.* Toll-like receptor 2-dependent inhibition of macrophage class II MHC expression and antigen processing by 19 kD lipoprotein of *Mycobacterium tuberculosis*. *J. Immunol.* **167**, 910–918 (2001).
20. Brightbill, H.D. *et al.* Host defense mechanisms triggered by microbial lipoproteins through toll-like receptors. *Science* **285**, 732–736 (1999).
21. Takeuchi, O. *et al.* Cutting edge: role of Toll-like receptor 1 in mediating immune response to microbial lipoproteins. *J. Immunol.* **169**, 10–14 (2002).
22. Nigou, J. *et al.* Mannan chain length controls lipoglycans signaling via and binding to TLR2. *J. Immunol.* **180**, 6696–6702 (2008).
23. Pecora, N.D., Gehring, A.J., Canaday, D.H., Boom, W.H. & Harding, C.V. *Mycobacterium tuberculosis* LprA is a lipoprotein agonist of TLR2 that regulates innate immunity and APC function. *J. Immunol.* **177**, 422–429 (2006).
24. Gehring, A.J., Dobos, K.M., Belisle, J.T., Harding, C.V. & Boom, W.H. *Mycobacterium tuberculosis* LprG (Rv1411c): A novel TLR-2 ligand that inhibits human macrophage class II MHC antigen processing. *J. Immunol.* **173**, 2660–2668 (2004).
25. Jung, S.B. *et al.* The mycobacterial 38-kilodalton glycolipoprotein antigen activates the mitogen-activated protein kinase pathway and release of proinflammatory cytokines through Toll-like receptors 2 and 4 in human monocytes. *Infect. Immun.* **74**, 2686–2696 (2006).
26. Jin, M.S. *et al.* Crystal structure of the TLR1–TLR2 heterodimer induced by binding of a tri-acylated lipopeptide. *Cell* **130**, 1071–1082 (2007).
27. Rezwan, M., Grau, T., Tschumi, A. & Sander, P. Lipoprotein synthesis in mycobacteria. *Microbiology* **153**, 652–658 (2007).
28. Nigou, J., Gilleron, M. & Puzo, G. Lipoarabinomannans: characterization of the multiacylated forms of the phosphatidyl-myo-inositol anchor by NMR spectroscopy. *Biochem. J.* **337**, 453–460 (1999).
29. Bigi, F. *et al.* The gene encoding P27 lipoprotein and a putative antibiotic-resistance gene form an operon in *Mycobacterium tuberculosis* and *Mycobacterium bovis*. *Microbiology* **146**, 1011–1018 (2000).
30. Gilleron, M., Nigou, J., Nicolle, D., Quesniaux, V. & Puzo, G. The acylation state of mycobacterial lipomannans modulates innate immunity response through toll-like receptor 2. *Chem. Biol.* **13**, 39–47 (2006).
31. Kang, J.Y. *et al.* Recognition of lipopeptide patterns by Toll-like receptor 2-Toll-like receptor 6 heterodimer. *Immunity* **31**, 873–884 (2009).
32. Berg, S., Kaur, D., Jackson, M. & Brennan, P.J. The glycosyltransferases of *Mycobacterium tuberculosis*—roles in the synthesis of arabinogalactan, lipoarabinomannan, and other glycoconjugates. *Glycobiology* **17**, 35R–56R (2007).
33. Finberg, R.W., Re, F., Popova, L., Golenbock, D.T. & Kurt-Jones, E.A. Cell activation by Toll-like receptors: role of LBP and CD14. *J. Endotoxin Res.* **10**, 413–418 (2004).
34. Jiang, Z. *et al.* CD14 is required for MyD88-independent LPS signaling. *Nat. Immunol.* **6**, 565–570 (2005).

ONLINE METHODS

Mammalian cell culture and cytokine ELISAs. Bone marrow was obtained from TLR1$^{-/-}$ TLR2$^{-/-}$ TLR6$^{-/-}$ CD14$^{-/-}$, wild-type C57BL/6 and F2 hybrids of C57BL/6 and 129sv mice (see **Supplementary Methods**). Bone marrow cells were cultured for 7–12 d in standard medium supplemented with 25% (v/v) LADMAC cell-conditioned medium[35]. HEK 293 cells stably expressing TLR2-YFP (HEK293.TLR2)[36,37], HEK293.pcDNA3 cells transfected with the empty vector. HEK293.TLR2-CD14 cells (Invivogen 293-htlr2cd14) or bone marrow–derived macrophages were incubated for 12–16 h with or without TLR2 agonist. IL-8 and TNFα were quantified by ELISA. Synthetic TLR agonists included FSL-1 (Invivogen, tlrl-fsl), Pam_3CSK_4 (Invivogen, tlrl-pms) and CpG-B ODN 1826 (5′-TCCATGACGTTCCTGACGTT-3′ lot C44-05225-q1a) provided by Coley Pharmaceutical Group.

Cloning and expression of hexahistidine-tagged proteins. LprA and NA-LprA were cloned previously[23]. LprG was amplified from MTB H37Rv genomic DNA by PCR, a nonacylated variant of LprG was cloned by using a 5′ primer that excluded the signal sequence and changed the acylated cysteine to a methionine, and site-directed mutagenesis was performed (**Supplementary Methods**). For expression in *M. smegmatis*, constructs were digested with NdeI and HindIII (NEB) and ligated into the shuttle vector pVV16 (provided by J. Belisle, Colorado State University) behind the constitutively active *hsp60* promoter and in frame with a C-terminal hexahistidine tag. For expression in *E. coli* Rosetta cells (EMD), constructs were digested with NdeI and HindIII and ligated with the expression plasmid pET-22b(+) (Novagen) (removing the pelB leader sequence), placing the coding sequence behind the IPTG-inducible T7 promoter and in frame with a C-terminal hexahistidine tag. Conditions for expression of recombinant lipoproteins in *M. smegmatis* and *E. coli* are described in **Supplementary Methods**.

Purification of hexahistidine-tagged proteins. Bacteria were disrupted mechanically by four passages through a French press (2,000 psi). Insoluble material was removed from the lysate by ultracentrifugation at 100,000*g* for 1 h at 4 °C. Recombinant lipoproteins were purified from the supernatant by nickel-affinity and anion-exchange chromatography (material eluted at 50–200 mM NaCl was used for all experiments, see **Supplementary Methods**). Protein purity was verified by SDS-PAGE with silver stain and anti-hexahistidine western blot. Both acylated and nonacylated forms were readily soluble in aqueous buffers.

Mycobacterial lysates and charging of *E. coli*–derived proteins. Lysates of *M. smegmatis* or MTB H37Ra were prepared by sonication (**Supplementary Methods**). MTB H37Rv lysates were obtained from K. Dobos-Elder and J. Belisle (Colorado State University, under NIH contract HHSN266200400091C, N01-AI-40091). For charging of *E. coli*–expressed proteins, 300–500 µg of protein (purified by nickel-affinity and anion-exchange chromatography) was incubated with mycobacterial lysate for 4 h at 37 °C. Insoluble material was pelleted, and charged proteins were repurified by nickel-affinity and anion-exchange chromatography. For charging with purified glycolipids, proteins similarly purified from *E. coli* (100 µg) were incubated for 3 h at 37 °C with 0.5–50 µg of purified glycolipid (Invivogen; LAM-MS, LM-MS) in a total volume of 100 µl and repurified.

SDS-PAGE and visualization of purified proteins and glycolipids. Gels (13% (w/v) acrylamide) were cast and run using a Tris-HCl buffer system. Proteins were visualized with Silver Stain Plus (BioRad). Carbohydrates (including glycolipids) were visualized with Pro-Q Emerald 300 (Molecular Probes) following periodate oxidation. Mycobacterial proteins and glycolipids were also visualized by Western blot analysis with rabbit polyclonal anti-BCG antibody (DAKO). To compare the amount of LAM normalized to the amount of hexahistidine-tagged protein, two-color Western blot analysis was performed using the Odyssey Western Analysis System (LI-COR). Materials were detected by sequential incubations with mouse monoclonal anti-LAM (CS-35, US National Institute of Allergy and Infectious Diseases HHSN266200400091C contract, Colorado State University), polyclonal goat anti-mouse (LI-COR), mouse monoclonal anti-hexahistidine (Santa Cruz) and donkey anti-mouse (LI-COR) antibodies.

MS and identification of ligands of NA-LprG. For nanospray MS, 12 µg of purified protein was denatured in 500 µL methanol, and 10 µL of methanol eluate was analyzed by negative-mode electrospray ionization MS (LCQ Advantage, Thermo Finnigan). For LC-MS analysis, 4.3 nmol of each protein was analyzed on a monochrome diol column (46 mm × 250 mm, 3 µm; Varian Inc.) coupled on line to a LXQ 2 dimensional ion-trap mass spectrometer (Thermo Finnigan) equipped with an electrospray ionization source (**Supplementary Methods**).

Expression and purification of NA-LprG for crystallography studies. Analysis of MTB LprG by secondary structure prediction and SignalP predicted the N-terminal 35 residues to be a signal peptide and the C-terminal 5 residues to be disordered, so these residues were excluded from LprG coding region (**Supplementary Methods**). The resulting sequence was ligated into pET30b (Novagen) and transformed into BL21 (DE3) (Novagen) and autotroph *E. coli* B834 (DE3) (Novagen). Cells were disrupted, and the lysate was centrifuged at 10,000*g* for 60 min. LprG was purified by nickel-affinity chromatography and concentrated to 10 mg ml^{-1} for crystallization. NA-LprG crystals were obtained from 0.1 M sodium acetate trihydrate, pH 4.5, 25% (w/v) PEG3350. SeMet crystals were obtained from 0.1 M bis-Tris, pH 6.5, 25% (w/v) PEG3350. NA-LprG V91W crystals were produced in 0.2 M ammonium sulfate, 0.1 M sodium acetate tri-hydrate, pH 4.6, 30% (w/v) PEG monomethyl ether 2000. Cocrystallization with glycolipid produced crystals in the NA-LprG crystal condition.

Determination of LprG structures. One three-wavelength MAD dataset was collected to resolution of 1.9 Å (**Table 1**) at the Lawrence Livermore National Laboratory. Native NA-LprG was collected to resolution of 2 Å on a Raxis IV++ detector in lab (Rigaku). The data was processed with HKL2000 (ref. 38). The space group of SeMet NA-LprG was *C*2, and cell dimensions were $a = 95.5$ Å, $b = 72.3$ Å, $c = 62.3$ Å, $\alpha = 90°$, $\beta = 106.8°$, $\gamma = 90°$. Native NA-LprG also crystallized in *C*2 with similar cell dimensions (**Table 1**). Both had two molecules in an asymmetric unit (molecules A and B). Phases were determined using the SeMet MAD NA-LprG dataset. An initial model from AUTOSHARP[39] was refined with CCP4 REFMAC[40] and PHENIX[41] and built with Coot[42]. The final R_{work} and R_{free} were 22.9% and 27.8%, respectively. Molecule A is used to represent the structure of NA-LprG in the text.

The crystal structure data for NA-LprG with PIM bound was collected at beamline 23ID-B in Argonne National Lab to a resolution of 1.8 Å (**Table 1**). HKL2000 was used for data processing. Apo NA-LprG structure was the template for rigid-body refinement with PHENIX. The structure of the ligand was built and refined with Coot, CCP4 and PHENIX. The final R_{work} and R_{free} were 22.2% and 25.8%, respectively. NA-LprG V91W crystal data were collected at 19ID in Argonne National Lab. High-resolution data were collected to 1.85 Å (**Table 1**) and processed with HKL2000 in $P2_1$ space group with cell dimensions $a = 39.7$ Å, $b = 56.0$ Å, $c = 96.9$ Å, $\alpha = 90°$, $\beta = 99.6°$, $\gamma = 90°$). Structure solution was obtained by molecular replacement[40]. The structure of NA-LprG V91W was refined with PHENIX and built with Coot. The final R_{work} and R_{free} were 21.5% and 25.3%, respectively. Protein structure graphics were produced using Chimera[43]. The predicted structure of LprA was generated using CPH model v2.0 (http://www.cbs.dtu.dk/services/CPHmodels/) and MODELLER (http://www.salilab.org/modeller). Cavity sizes of proteins were calculated using the CastP program with spherical probe of 1.4 Å (ref. 44).

35. Sklar, M.D., Tereba, A., Chen, B.D. & Walker, W.S. Transformation of mouse bone marrow cells by transfection with a human oncogene related to c-myc is associated with the endogenous production of macrophage colony stimulating factor 1. *J. Cell. Physiol.* **125**, 403–412 (1985).
36. Flo, T.H. *et al.* Involvement of toll-like receptor (TLR) 2 and TLR4 in cell activation by mannuronic acid polymers. *J. Biol. Chem.* **277**, 35489–35495 (2002).
37. Latz, E. *et al.* Lipopolysaccharide rapidly traffics to and from the Golgi apparatus with the toll-like receptor 4-MD-2–CD14 complex in a process that is distinct from the initiation of signal transduction. *J. Biol. Chem.* **277**, 47834–47843 (2002).
38. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
39. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364**, 215–230 (2007).
40. Murshudov, G.N., Vagin, A.A. & Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255 (1997).
41. Adams, P.D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1948–1954 (2002).
42. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
43. Pettersen, E.F. *et al.* UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
44. Dundas, J. *et al.* CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res.* **34**, W116–8 (2006).

doi:10.1038/nsmb.1869

APPENDIX B

THE STRUCTURAL STUDY OF TRICLOSAN DERIVATIVES WITH MODIFICATION AT THE 2'-POSITION

*Reprinted with permission from "Synthesis and biological activity of diaryl ether inhibitors of malarial enoyl acyl carrier protein reductase. Part 2: 2'-substituted triclosan derivatives" by Freundlich JS, Yu M, Lucumi E, Kuo M, Tsai HC, Valderramos JC, Karagyozov L, Jacobs WR Jr, Schiehser GA, Fidock DA, Jacobus DP, Sacchettini JC, 2006. Bioorganic & Medicinal Chemistry Letters, 16(8), 2163-2169, Copyright 2006 by Elsevier Ltd.

Available online at www.sciencedirect.com
SCIENCE DIRECT
Bioorganic & Medicinal Chemistry Letters
ELSEVIER
Bioorganic & Medicinal Chemistry Letters 16 (2006) 2163–2169

Synthesis and biological activity of diaryl ether inhibitors of malarial enoyl acyl carrier protein reductase. Part 2: 2′-Substituted triclosan derivatives

Joel S. Freundlich,[a,*] Min Yu,[b] Edinson Lucumi,[d] Mack Kuo,[d] Han-Chun Tsai,[d] Juan-Carlos Valderramos,[b] Luchezar Karagyozov,[b] William R. Jacobs, Jr.,[b,c] Guy A. Schiehser,[a] David A. Fidock,[b] David P. Jacobus[a] and James C. Sacchettini[d]

[a]*Department of Medicinal Chemistry, Jacobus Pharmaceutical Company, 37 Cleveland Lane, Princeton, NJ 08540, USA*
[b]*Department of Microbiology and Immunology, Albert Einstein College of Medicine, Bronx, NY 10461, USA*
[c]*Howard Hughes Medical Institute, Albert Einstein College of Medicine, Bronx, NY 10461, USA*
[d]*Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX 77843-2128, USA*

Received 26 October 2005; revised 13 January 2006; accepted 13 January 2006
Available online 8 February 2006

Abstract—2′-Substituted analogs of triclosan have been synthesized to target inhibition of the key malarial enzyme *Plasmodium falciparum* enoyl acyl carrier protein reductase (PfENR). Many of these compounds exhibit good potency ($EC_{50} < 500$ nM) against in vitro cultures of drug-resistant and drug-sensitive strains of the *P. falciparum* parasite and modest ($IC_{50} = 1–20$ μM) potency against purified PfENR enzyme. Compared to triclosan, this survey of 2′-substituted derivatives has afforded gains in excess of 20- and 30-fold versus the 3D7 and Dd2 strains of parasite, respectively.

Treatment of *Plasmodium falciparum* malaria has depended for decades on the use of the aminoquinoline chloroquine or the antifolate combination pyrimethamine-sulfadoxine. *P. falciparum* infection is the most widespread and the deadliest form of malaria—a disease that afflicts 300–600 million people worldwide resulting in 1–3 millions deaths per annum.[1] The occurrence and spread of drug-resistance necessitates new, highly potent antimalarials that are not encumbered by existing resistance mechanisms and are orally biovailable, affordable, and safe.

The inhibition of fatty acid synthesis in *P. falciparum* has been a recent focus of drug discovery efforts to find novel antimalarials.[2] This synthetic pathway is believed to be crucial to parasite survival as a result of its key role in membrane construction and energy production. Synthesis of fatty acids is regulated within the apicoplast—an organelle derived from a cyanobacterial endosymbiont.[3] Given the prokaryotic nature of the apicoplast, its metabolic machinery differs significantly from that of mammalian cells.[4] For example, higher eukaryotes and yeast utilize fatty acid biosynthetic machinery residing on a single multifunctional polypeptide (FAS-I), whereas fatty acid synthesis in *Plasmodium* is reliant on a dissociative process that utilizes a set of distinct enzymes composing a FAS-II pathway. This observation hints at selectivity with regard to the human host in a therapeutic strategy.

The enoyl reductase enzyme (PfENR) is responsible for the final step in fatty acid synthesis: NADH-dependent reduction of *trans*-2-enoyl-ACP to acyl-ACP (ACP = acyl carrier protein). Triclosan (Fig. 1) is a known

Triclosan

Figure 1. Chemical structure of triclosan with demarcation of key B-ring sites.

Keywords: Antimalarial; Diaryl ether; Phenol.
* Corresponding author. Tel.: +1 6099217447; fax: +1 6097991176; e-mail: J_Freundlich@jacobuspharm.com

0960-894X/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2006.01.051

inhibitor of PfENR, as determined with native enzyme purified from *P. falciparum* or recombinant enzyme expressed in *Escherichia coli*.[5,6] Furthermore, triclosan is potent in the low micromolar range against *P. falciparum*[5,7–10] and is reported to effectively treat *P. berghei* in vivo infections in mice.[5] A direct relation between triclosan binding to and inhibition of PfENR in vitro and its whole cell properties is supported by the demonstration that triclosan also inhibits fatty acid synthesis with cultured parasites or parasite extracts.[5] Given triclosan's abundance (U.S. annual production $> 1 \times 10^6$ lbs),[11] its safety as demonstrated by its widespread use in personal care and household products, and the absence of a lipid synthesis inhibitor in the antimalarial arsenal, we and others have chosen to pursue the discovery of a triclosan-derived therapeutic for malaria.[6,8,10,12–16]

This work recently culminated in the X-ray crystal structure of triclosan and NAD^+ co-factor bound to PfENR.[8] The structure demonstrated the critical binding element of triclosan to be the phenol moiety, while suggesting vectors off the triclosan A and B rings where additional functionality could be placed to enhance binding affinity. We have pursued a strategy to build off these aromatic ring positions to enhance enzymatic and anti-parasitic activity. A significant gain in potency could facilitate replacement or removal of the phenol, due to its metabolic liability.[17] Previously, we reported efforts to explore SAR at the 4′-position which resulted in minor improvements in enzyme and anti-parasite potency.[16] This report details our further exploration of the triclosan B-ring, focusing specifically on the 2′-position.

Examination of the triclosan:co-factor:PfENR structure showed the 2′-chloro is pointed toward the pyrophosphate portion of NAD^+ and is approximately 4 Å from two of the negatively charged oxygens (Fig. 2). We hypothesized that a properly placed positively charged amine off the 2′-position could favorably interact with one or both of these oxygens. In addition, a hydrophobic substituent on the amine could participate in favorable interactions with proximal enzyme residues such as Ala-319 and Arg-318. Compared to triclosan, 2′-analogs with these features could have enhanced interactions with both the bound co-factor and enzyme, thus conferring greater binding affinity for the co-factor:enzyme complex and potentially enhanced anti-parasitic activity. Amino derivatives off the 2′-position could also have improved physiochemical properties.

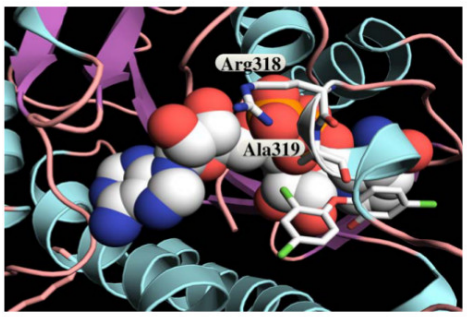


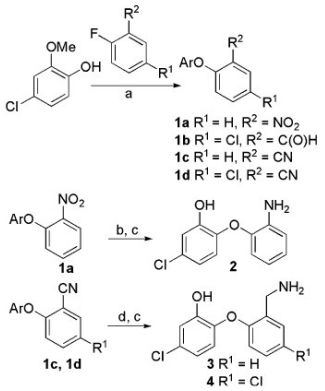
Figure 2. X-ray crystal structure of triclosan (stick drawing) bound to PfENR (ribbon and tube with key residues in stick format) in the presence of NAD^+ (space-fill). The 2′-chloro is shown pointing towards Ala-319 and Arg-318.

The synthesis of triclosan analogs began with diaryl ether anisoles **1a**–**d**, prepared from commercially available materials via nucleophilic aromatic substitution (Scheme 1). Aniline **2** was synthesized via hydrogenation of nitroarene **1a** followed by boron(III) bromide-mediated cleavage of the methyl ether. Benzonitriles **1c** and **1d** were converted to the 2′-aminomethyl compounds **3** and **4** via reduction with lithium aluminum hydride and then methyl ether cleavage. In all cases, final compounds were characterized by ^1H NMR and LC–MS.

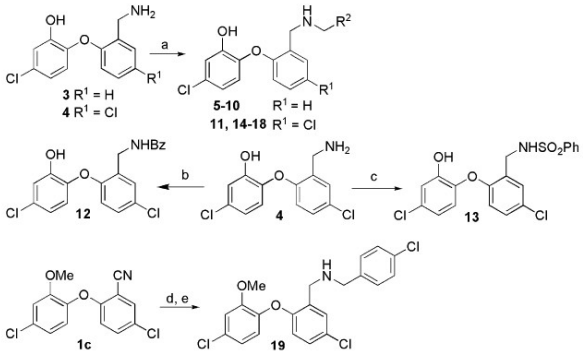
As shown in Scheme 2, reductive amination, utilizing solid-phase reagents/scavengers[18] and various aldehydes, allowed the functionalization of aminomethyl derivatives **3** and **4** to provide **5**–**10**, **11**, and **14**–**18**. Interestingly, the reductive aminations proceeded more cleanly (i.e., less dialkylation) with the corresponding phenol than with the anisole (free base or hydrochloride salt). Compound **4** was also coupled with benzoic acid to afford amide **12** and sulfonylated with benzenesulfonyl chloride to yield sulfonamide **13**.

Anisole **1c** could be converted to benzylic amine **19** via nitrile reduction with lithium aluminum hydride followed by reductive amination using the protocol of Bhattacharyya.[19]

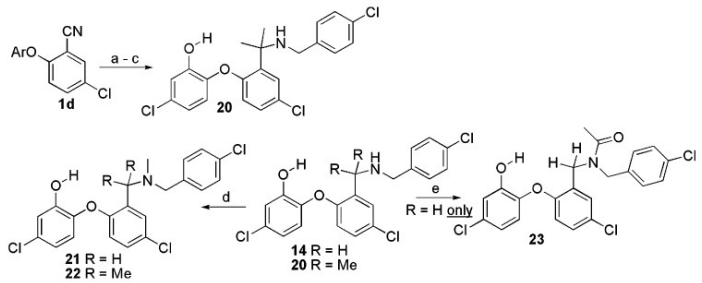
Analogs of **14** were prepared as shown in Scheme 3. Benzonitrile **1d** was transformed to α,α-dimethylamino derivative **20** via a three-step sequence, featuring reduc-



Scheme 1. Reagents and conditions: (a) K_2CO_3, DMSO, Δ; (b) Ra-Ni, H_2, EtOH/EtOAc; (c) BBr_3, DCM, 0 °C to rt; (d) $LiAlH_4$, THF. Ar = 4-chloro-2-methoxyphenyl.



Scheme 2. Reagents and conditions: (a) i—MP-Triacetoxyborohydride, RCHO, THF; ii—PS-PhCHO; iii—MP-TsOH; (b) $PhCO_2H$, ACTU, PS-HOBt, DIEA, DMF; (c) $PhSO_2Cl$, pyridine, DCM/THF; (d) $LiAlH_4$, THF; (e) i—$Ti(O\text{-}i\text{-}Pr)_4$, $p\text{-}ClC_6H_4CHO$, THF/EtOH; ii—$NaBH_4$, EtOH.

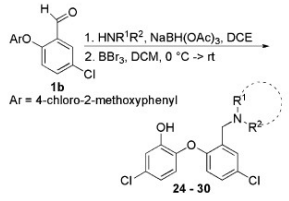


Scheme 3. Reagents and conditions: (a) $CeCl_3$, MeLi, THF, −78 °C to rt; (b) BBr_3, DCM, 0 °C to rt; (c) i—$Ti(O\text{-}i\text{-}Pr)_4$, $p\text{-}ClC_6H_4CHO$, THF/EtOH; ii—$NaBH_4$, EtOH; (d) $Ti(O\text{-}i\text{-}Pr)_4$, $NaBH_4$, $(HCHO)_n$, EtOH; (e) Ac_2O, DCM. Ar = 4-chloro-2-methoxyphenyl.

tive alkylation of the nitrile using an excess of in situ prepared methylorganocerium reagent.[20] N-methylation of **14** and **20** was achieved via a $Ti(O\text{-}i\text{-}Pr)_4/NaBH_4$ mediated reaction with paraformaldehyde to provide **21** and **22**, respectively.[21] N-acetylation, however, was only attainable with **16** to afford acetamide **23**.

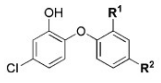
Reductive aminations to afford tertiary amines were initiated from benzaldehyde **1b**. In contrast to reductive aminations with aminomethyl analogs **3** and **4**, **1b** and various amines were observed to react most smoothly with regard to a reductive amination pathway in the presence of two equivalents of sodium triacetoxyborohydride in 1,2-dichloroethane (Eq. 1).[22] This synthetic protocol, followed by boron(III) bromide-promoted methyl ether cleavage, afforded tertiary amines **24**–**30**.

Final compounds **2**–**30** were tested in two assay systems to determine their inhibition of PfENR enzymatic activity and inhibition of parasite whole cell growth. For both assays, the reported inhibitory concentrations are tabulated as means ± standard error, taking into account three independent determinations performed in duplicate. Two parasite strains were utilized: 3D7, which is drug-sensitive, and Dd2, which is resistant to chloroquine and pyrimethamine-sulfadoxine. Both systems were previously described in detail.[8,16] To date, fourteen independent determinations of the IC_{50} of triclosan have been made in the biochemical screen with purified PfENR, resulting in a mean of 73 nM with a standard error of 21 nM. For the parasite whole-cell assay, 46 independent runs of triclosan have been conducted. Against the 3D7 strain, the EC_{50} was determined to have a mean value of 2.9 μM and a standard error of 0.2 μM versus the Dd2 strain, the EC_{50} mean value for triclosan was 3.8 μM with a standard error of 0.2 μM.



(1)

Table 1. Inhibitory properties of selected 2′-substituted derivatives



Compound	R^1	R^2	EC_{50}[a] 3D7/Dd2 (μM)	PfENR[a] IC_{50} (μM)
Triclosan	Cl	Cl	2.9 ± 0.2/3.8 ± 0.2	0.073 ± 0.021
2	NH_2	H	83 ± 15/140 ± 20	7 ± 2
3	CH_2NH_2	H	9.7 ± 1.5/7.2 ± 0.8	>50

[a] Values reported as means ± standard error.

The first set of 2′-derivatives prepared is shown in Table 1 where the 4′-position was chosen to be hydrogen. Although aminomethyl **3** is much less potent than aniline **2** against the purified enzyme, it is significantly more efficacious against the cultured parasites.

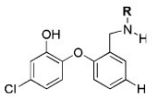
In keeping with our design hypothesis, we examined the effect of placing hydrophobic groups off the amine terminus of **3** in order to improve its enzyme and anti-parasitic potencies over those of triclosan. As shown in Table 2, compounds **5**–**10** achieve approximately an order of magnitude increase in potency over **3** in the parasite assay. The 1-naphthyl **9**, phenylpropyl **8**, and *p*-Cl-benzyl **6** derivatives are the most potent members of this series against the parasite. In general, the compounds also display an improvement in enzyme activity over **3**, although their micromolar level IC_{50} values against PfENR appear incongruent with their nanomolar level anti-parasitic potencies. It is interesting to note that of the 2′-substituted derivatives prepared to date, phenethyl analog **7** is the most potent against purified enzyme (IC_{50} = 2.5 μM).

Subsequently, triclosan derivatives with a 4′-chloro were prepared in an effort to improve enzyme and anti-parasitic potency while adhering more strictly to the triclosan scaffold (Table 3). While **4** is observed to be more potent than **3** against cultured parasites, it fails to offer an observable improvement in the enzyme assay. Surprisingly, introduction of the 4′-chloro group does not consistently improve the performance of these analogs in both assay systems. Phenylamide **12** and phenylsulfonamide **13** derivatives are less active than **11** versus the enzyme and parasite. Whereas utilization of polar moieties was not tolerated, proper placement of hydrophobic functionality off the phenyl ring of **11** generally afforded more potent anti-parasitic compounds while leading to reduced enzyme inhibition. For example, the 2-naphthyl derivative **17** exhibits EC_{50} values of 210 and 140 nM against the two parasite strains, while the 5-(2,3-dihydrobenzofuranyl) analog **18** displays EC_{50} values of 180 and 110 nM against 3D7 and Dd2. Compound **18** exhibits the most potent activity against Dd2 in the 2′-substituted class.

Analogs of **6** and **14**, containing a *p*-chlorobenzyl-2′-aminomethyl substituent, were investigated further and the biological data are shown in Table 4. The importance of the phenol is underscored by the observation that anisole **19** is much less potent than phenol **6** against both the enzyme and parasite. Comparison of the data for **6** and **14** shows that alteration of the 4′-substituent from hydrogen to chloride confers slightly more potency against PfENR, while not significantly affecting anti-parasitic activity. Interestingly, α,α-dimethylation of the 2′-benzylic carbon (cf. **20**) slightly reduces the parasite potency, while having no effect in the enzyme assay. N-methylation (cf. **21**) leads to a slight gain in the parasite assay and a loss in enzyme inhibition. The combination of these two structural alterations in **22** leads to increased anti-parasitic potency against the 3D7 strain but not the Dd2 strain, with a slight loss in enzyme activity. **22** exhibits the most potent activity in the 2′-substituted chemotype against the 3D7 strain of parasite (EC_{50} = 140 nM). N-acetylation to afford **23** produces losses in efficacy in both assay systems.

In light of the enhanced anti-parasitic activity of **21** due to its N-methylation, a series of 2′-*N*-methylamines was synthesized and evaluated in both screening assays (Table 5). Mono- and dimethylamino analogs **24** and **25** show enhanced activity compared to that of parent amino **4** in both assays. The two *N*-methyl substituents were next constrained in a ring system. Cyclic analogs (**26**–**27**) display slight losses in parasite activity and are inactive against PfENR. The N4-substituent of piperazines **28**–**30** modulates potency against both cultured parasites

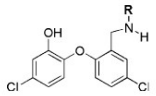
Table 2. Effect of secondary amine substituent with 4′-hydrogen

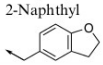


Compound	R	EC_{50}[a] 3D7/Dd2 (μM)	PfENR IC_{50}[a] (μM)
3	H	9.7 ± 1.5/7.2 ± 0.8	>50
5	CH_2Ph	0.49 ± 0.02/0.42 ± 0.03	10 ± 3
6	$CH_2(p\text{-}ClC_6H_4)$	0.33 ± 0.04/0.30 ± 0.04	35 ± 4
7	$(CH_2)_2Ph$	0.78 ± 0.05/0.75 ± 0.07	2.5 ± 0.9
8	$(CH_2)_3Ph$	0.32 ± 0.08/0.39 ± 0.08	19 ± 5
9	1-Naphthyl	0.30 ± 0.02/0.37 ± 0.07	11 ± 4
10	2-Naphthyl	0.45 ± 0.05/0.33 ± 0.09	46 ± 14

[a] Values reported as means ± standard error.

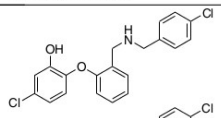
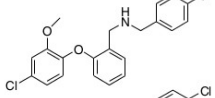
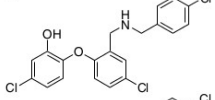
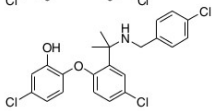
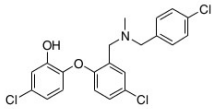
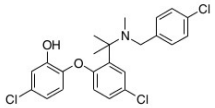
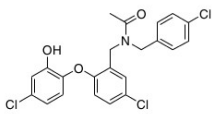
Table 3. Effect of secondary amine substituent with 4′-chloride



Compound	R	EC_{50}[a] 3D7/Dd2 (μM)	PfENR[a] IC_{50} (μM)
4	H	2.7 ± 0.1/2.0 ± 0.5	>50
11	CH_2Ph	0.77 ± 0.05/0.83 ± 0.01	3.3 ± 1.5
12	C(O)Ph	8.6 ± 1.8/12 ± 1	18 ± 8
13	SO_2Ph	11 ± 1/17 ± 1	15 ± 7
14	$CH_2(p\text{-}ClC_6H_4)$	0.32 ± 0.08/0.27 ± 0.10	7.2 ± 1.4
15	$CH_2(p\text{-}PhC_6H_4)$	0.23 ± 0.06/0.15 ± 0.04	35 ± 15
16	1-Naphthyl	2.6 ± 0.1/2.2 ± 0.1	>50
17	2-Naphthyl	0.21 ± 0.06/0.14 ± 0.03	26 ± 7
18		0.18 ± 0.03/0.11 ± 0.01	9 ± 3

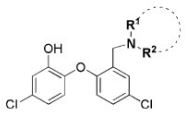
[a] Values reported as means ± standard error.

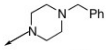
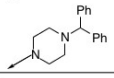
Table 4. Inhibitory properties of *p*-Cl-benzylamine derivatives

Compound	Chemical structure	EC_{50}[a] 3D7/Dd2 (μM)	PfENR[a] IC_{50} (μM)
6		0.33 ± 0.04/0.30 ± 0.04	35 ± 4
19		19 ± 2/24 ± 1	>50
14		0.32 ± 0.08/0.27 ± 0.10	7.2 ± 1.4
20		0.41 ± 0.03/0.35 ± 0.02	7.2 ± 1.6
21		0.25 ± 0.06/0.19 ± 0.03	22 ± 2
22		0.14 ± 0.01/ 0.38 ± 0.02	27 ± 4
23		3.5 ± 0.2/5.9 ± 0.1	28 ± 3

[a] Values reported as means ± standard error.

Table 5. Inhibitory properties of selected amines



Compound	NR^1R^2	EC_{50}[a] 3D7/Dd2 (μM)	PfENR[a] IC_{50} (μM)
24	N(H)Me	0.58 ± 0.09/0.47 ± 0.13	27 ± 9
25	NMe_2	0.18 ± 0.01/0.22 ± 0.01	19 ± 5
26		0.50 ± 0.03/0.45 ± 0.03	>50
27		0.73 ± 0.21/0.48 ± 0.26	>50
28		14 ± 2/15 ± 3	6.1 ± 1.0
29		0.63 ± 0.05/0.62 ± 0.03	11 ± 2
30		2.1 ± 0.1/1.4 ± 0.1	>50

[a] Values reported as mean ± standard error.

and purified enzyme. Compared to piperidine **27**, significant gains in enzymatic potency against the enzyme are realized with **28** (IC_{50} = 6.1 μM). The corresponding anti-parasitic activities of these piperazines do not improve significantly upon that of **27**.

In conclusion, a series of 2′-substituted triclosan derivatives has been prepared and assayed for inhibition of PfENR and intracellular parasite growth. The effort provided a number of compounds that are significantly more potent than triclosan against the cultured parasites. 5-(2,3-dihydrobenzofuranyl) analog **18** was 16 and 34 times more active than triclosan against the 3D7 and Dd2 strains, respectively. Methylated analog **22** displayed anti-parasitic efficacy 21 and 10 times greater than triclosan against 3D7 and Dd2, respectively. Studies are now required to assess whether these compounds are cytotoxic or cytostatic. Preliminary indications that this compound class is cytotoxic come from the study of Waller et al.,[9] who exposed synchronized parasites at different stages to triclosan for 16 h periods. With ring stage parasites, susceptibility to triclosan was similar to that observed when parasites were exposed for the full 48 h of intracellular development, suggesting that the drug action was toxic and not static. Nevertheless, further experiments are required to rigorously address this, including 'washout' experiments and these will be performed with the more potent compounds from the triclosan series.

Given the micromolar activity of these and other 2′-analogs versus purified PfENR, it appears reasonable to suggest the possibility of off-target activity. Future efforts will seek to define the mode of action of these potent antimalarials while further optimizing their activity. These include studies on whether PfENR is the primary target (indeed, it may very well not be the intracellular target for these compounds) and parallel genetic investigations to assess whether this target is essential. This will be done in conjunction with probing other positions along the diaryl ether scaffold of triclosan.

Acknowledgment

This work has been supported by funding from the Medicines for Malaria Venture.

References and notes

1. Fidock, D. A.; Rosenthal, P. J.; Croft, S. L.; Brun, R.; Nwaka, S. *Nat. Rev. Drug Discov.* **2004**, *3*, 509.
2. Lu, J. Z.; Lee, P. J.; Water, N. C.; Prigge, S. T. *Comb. Chem. High Throughput Screening* **2005**, *8*, 15.
3. McFadden, G. I.; Reith, M. E.; Munholland, J.; Lang-Unnasch, N. *Nature* **1996**, *381*, 482.
4. Schweizer, E.; Hofmann, J. *Microbiol. Mol. Biol. Rev.* **2004**, *68*, 501.
5. Surolia, N.; Surolia, A. *Nat. Med.* **2001**, *7*, 167.
6. Kapoor, M.; Reddy, C. C.; Krishnasastry, M. V.; Surolia, N.; Surolia, A. *Biochem. J.* **2004**, *381*, 719.
7. McLeod, R.; Muench, S. P.; Rafferty, J. B.; Kyle, D. E.; Mui, E. J.; Kirisits, M. J.; Mack, D. G.; Roberts, C. W.; Samuel, B. U.; Lyons, R. E.; Dorris, M.; Milhous, W. K.; Rice, D. W. *Int. J. Parasitol.* **2001**, *31*, 109.
8. Perozzo, R.; Kuo, M.; bir Singh Sidhu, A. S.; Valiyaveettil, J. T.; Bittman, R.; Jacobs, W. R., Jr.; Fidock, D. A.; Sacchettini, J. C. *J. Biol. Chem.* **2002**, *277*, 13106.
9. Waller, R. F.; Ralph, S. A.; Reed, M. B.; Su, V.; Douglas, J. D.; Minnikin, D. E.; Cowman, A. F.; Besra, G. S.; McFadden, G. I. *Antimicrob. Agents Chemother.* **2003**, *47*, 297.

10. Kuo, M. R.; Morbidoni, H. R.; Alland, D.; Sneddon, S. F.; Gourlie, B. B.; Staveski, M. M.; Leonard, M.; Gregory, J. S.; Janjigian, A. D.; Yee, C.; Musser, J. M.; Kreiswirth, B.; Iwamoto, H.; Perozzo, R.; Jacobs, W. R., Jr.; Sacchettini, J. C.; Fidock, D. A. *J. Biol. Chem.* **2003**, *278*, 20851.
11. www.scorecard.com/chemical~profiles.
12. Surolia, N.; RamachandraRao, S. P.; Surolia, A. *BioEssays* **2002**, *24*, 192.
13. Rao, S. P. R.; Surolia, A.; Surolia, N. *Mol. Cell. Biochem.* **2003**, *253*, 55.
14. Kapoor, M.; Mukhi, P. L. S.; Surolia, N.; Suguna, K.; Surolia, A. *Biochem. J.* **2004**, *381*, 725.
15. Kapoor, M.; Gopalakrishnapai, J.; Surolia, N.; Surolia, A. *Biochem. J.* **2004**, *381*, 735.
16. Freundlich, J. S.; Anderson, J. W.; Sarantakis, D.; Shieh, H.-M.; Yu, M.; Valderramos, J.-C.; Lucumi, E.; Kuo, M.; Jacobs, W. R., Jr.; Fidock, D. A.; Schiehser, G. A.; Jacobus, D. P.; Sacchettini, J. C. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 5247.
17. Wang, L.-Q.; Falany, C. N.; James, M. O. *Drug Metab. Dispos.* **2004**, *32*, 1162.
18. Purchased from Argonaut Technologies, Incorporated.
19. Bhattacharyya, S.; Fan, L.; Vo, L.; Labadiie, J. *Comb. Chem. High Throughput Screening* **2000**, *3*, 117.
20. Ciganek, E. *J. Org. Chem.* **1992**, *57*, 4521.
21. Bhattacharyya, S. *Tetrahedron Lett.* **1994**, *35*, 2401.
22. Abdel-Magid, A. F.; Carson, K. G.; Harris, B. D.; Maryanoff, C. A.; Shah, R. D. *J. Org. Chem.* **1996**, *61*, 3849.

APPENDIX C

THE STRUCTURAL STUDY OF TRICLOSAN DERIVATIVES WITH MODIFICATION AT THE 5-POSITION

*Reprinted with permission from "X-ray structural analysis of Plasmodium falciparum enoyl acyl carrier protein reductase as a pathway toward the optimization of triclosan antimalarial efficacy." by Freundlich JS, Wang F, Tsai HC, Kuo M, Shieh HM, Anderson JW, Nkrumah LJ, Valderramos JC, Yu M, Kumar TR, Valderramos SG, Jacobs WR Jr, Schiehser GA, Jacobus DP, Fidock DA, Sacchettini JC, 2007. Journal of Biological Chemistry, 282(35), 25436-44, Copyright 2007 by The American Society for Biochemistry and Molecular Biology, Inc.

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 282, NO. 35, pp. 25436–25444, August 31, 2007

X-ray Structural Analysis of *Plasmodium falciparum* Enoyl Acyl Carrier Protein Reductase as a Pathway toward the Optimization of Triclosan Antimalarial Efficacy*[S]

Received for publication, March 1, 2007, and in revised form, June 5, 2007 Published, JBC Papers in Press, June 13, 2007, DOI 10.1074/jbc.M701813200

Joel S. Freundlich[‡1], Feng Wang[§], Han-Chun Tsai[§], Mack Kuo[§], Hong-Ming Shieh[‡], John W. Anderson[‡], Louis J. Nkrumah[¶], Juan-Carlos Valderramos[¶2], Min Yu[¶2], T. R. Santha Kumar[¶], Stephanie G. Valderramos[¶2], William R. Jacobs, Jr.[¶‖], Guy A. Schiehser[‡], David P. Jacobus[‡], David A. Fidock[¶2], and James C. Sacchettini[§3]

From the [‡]Department of Medicinal Chemistry, Jacobus Pharmaceutical Company, Princeton, New Jersey 08540, the [§]Department of Biochemistry and Biophysics, Texas A&M University, College Station, Texas 77843-2128, the [¶]Department of Microbiology and Immunology and the [‖]Howard Hughes Medical Institute, Albert Einstein College of Medicine, Bronx, New York 10461

The x-ray crystal structures of five triclosan analogs, in addition to that of the isoniazid-NAD adduct, are described in relation to their integral role in the design of potent inhibitors of the malarial enzyme *Plasmodium falciparum* enoyl acyl carrier protein reductase (PfENR). Many of the novel 5-substituted analogs exhibit low micromolar potency against *in vitro* cultures of drug-resistant and drug-sensitive strains of the *P. falciparum* parasite and inhibit purified PfENR enzyme with IC_{50} values of <200 nM. This study has significantly expanded the knowledge base with regard to the structure-activity relationship of triclosan while affording gains against cultured parasites and purified PfENR enzyme. In contrast to a recent report in the literature, these results demonstrate the ability to improve the *in vitro* potency of triclosan significantly by replacing the suboptimal 5-chloro group with larger hydrophobic moieties. The biological and x-ray crystallographic data thus demonstrate the flexibility of the active site and point to future rounds of optimization to improve compound potency against purified enzyme and intracellular *Plasmodium* parasites.

Malaria is a disease of worldwide significance, which is responsible for over one million deaths annually, mainly in children under the age of 5 years (1). *Plasmodium falciparum* infection is the most virulent form of malaria, accounting for the vast majority of deaths and >500 million infections yearly. The occurrence and spread of resistance to traditional drugs, such as chloroquine and sulfadoxine-pyrimethamine necessitate new, highly potent antimalarials that are orally bioavailable, affordable, safe, and unencumbered by existing resistance mechanisms (2).

Inhibition of type II fatty acid biosynthesis (FAS-II)[4] appears to hold significant promise in devising novel antimalarials that meet these criteria (3–5). This dissociative FAS-II pathway, which occurs in the plastid-like apicoplast organelle (6), is composed of four separate enzymes. As such, it is fundamentally different from the associative FAS-I multifunctional polypeptide present in mammalian cells. Fatty acid synthesis by the parasite has been postulated to be crucial to membrane construction and energy production. The final reaction in the FAS-II pathway is catalyzed by the enoyl acyl carrier protein reductase enzyme (PfENR in *P. falciparum*, also known as PfFabI), which mediates the NADH-dependent reduction of *trans*-2-enoyl-ACP (acyl carrier protein) to acyl-ACP.

Triclosan (Fig. 1) is an uncompetitive inhibitor of purified PfENR (7, 8) and demonstrates potency against *P. falciparum* parasites cultured *in vitro* (7, 9, 10–12) as well as *Plasmodium berghei in vivo* infections in mice (7). A possible correlation among these biological activities is supported by the observation that triclosan produces a dose-dependent inhibition of [^{14}C]acetate incorporation into fatty acids by cultured *P. falciparum* parasites and [^{14}C]malonyl-CoA incorporation into parasite extracts freed from host red blood cells (7) and also inhibits a *P. falciparum* FAS-II *in vitro* reconstituted system (13). A requirement for fatty acids was demonstrated by the finding that *P. falciparum in vitro* growth was dependent on the presence of fatty acids in the medium, with a combination of C14, C16, and/or C18 fatty acids being sufficient to replace Albumax (14). Given that current antimalarial treatments do not feature a lipid synthesis inhibitor, we and others have explored triclosan analogs as potential therapeutics (10, 15–17). Of note, triclosan has displayed a satisfactory safety profile when administered either topically in humans or systemically in mice (7, 18). This is especially important in light of evidence suggesting that mammalian cells harbor a mitochondrial FAS-II system, producing a lipoic acid precursor, which could have raised concerns of host toxicity (19–22).

* This work was supported by funding from the Medicines for Malaria Venture and National Institutes of Health Grants AI060342 and AI43268. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 2NQ8, 2FOI, 2OOS, 2OL4, 2OP0, and 2OP1) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Table S1 and Figs. S1–S7.

[1] To whom correspondence should be addressed: Dept. of Chemistry, Princeton University, Frick Chemical Laboratory, Princeton, NJ 08544. Tel.: 609-258-9804; Fax: 609-258-1980; E-mail: joelf@princeton.edu.

[2] Present address: Depts. of Microbiology and of Medicine, Columbia University, New York, NY 10032.

[3] Supported by Robert A. Welch Foundation Grant A-0015.

[4] The abbreviations used are: FAS-II, type II fatty acid biosynthesis; PfENR, *P. falciparum* enoyl acyl carrier protein reductase; INH, isoniazid.

Key to our triclosan analog program has been the x-ray crystal structure of triclosan and its co-factor bound to PfENR (10). The phenol moiety has been clearly demonstrated to be crucial to triclosan binding to both the enzyme and co-factor. However, *in vivo* conjugation of the phenol is a metabolic liability (23). The crystal structure was utilized to suggest vectors off the triclosan diaryl ether scaffold where functionality could be appended to gain energetically favorable interactions with both the enzyme and co-factor. Sufficient increases in potency *versus* both PfENR and cultured parasites could facilitate eventual replacement or excision of the phenol. A campaign to explore SAR at the 4′-position resulted in minor gains in enzyme and anti-parasite potency, potentially due to enhanced hydrogen-bonding interactions with proximal active site residues (15). Examination of analogs at the 2′-position afforded very potent antiparasitic agents ($EC_{50} < 200$ nM) that were only micromolar level inhibitors of PfENR activity (17).

Inspection of the PfENR·triclosan structure shows that the 5-chloro is in van der Waals contact with the side chains of Tyr-267, Pro-314, and Phe-368. We thus hypothesized that larger hydrophobic groups off the 5-position may better fill this enzyme pocket and confer greater potency against the purified enzyme and cultured parasites. More hydrophobic analogs of triclosan may also result in greater compound permeability to the four membranes surrounding the apicoplast (4, 24).

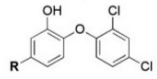
Two groups have recently published work describing 5-substituted triclosan analogs (16, 25). Sullivan *et al.* (25) reported 5-alkyl-substituted triclosan derivatives assayed against *Mycobacterium tuberculosis* enoyl acyl ACP reductase (InhA) that did not feature chlorine substitution on the B-ring. Chhibber *et al.* (16) described the synthesis and activity, *versus* both purified PfENR and cultured parasites, of triclosan analogs with primarily *hydrophilic* 5-substituents. These compounds were significantly less potent than triclosan in both assay systems. This led the investigators to propose that chloride is the ideal 5-moiety given its small size. Our studies, described below, do not support this finding and lead us to identify alternative 5-substituted analogs with enhanced potency.

EXPERIMENTAL PROCEDURES

Synthesis of 5-Substituted Triclosan Analogs—The compounds depicted in Table 1 were prepared via a route reliant on the coupling of A-ring and B-ring precursors through a nucleophilic aromatic substitution reaction, following protocols similar to those published previously (15, 17). In brief, A-ring phenol precursors were either purchased or prepared via the Suzuki-Miyaura reaction of 4-bromo-2-methoxyphenol with an arylboronic acid (26). Alternatively, the diaryl ether scaffold was constructed by the reaction of a 4-substituted-2-methoxyaryl fluoride with 2,4-dichlorophenol.

With the exception of **2**, benzaldehyde **A** was a key intermediate on the way to the 5-substituted alkyl derivatives shown in Table 2 (Scheme 1). Reaction of the aldehyde with an excess of the appropriate Grignard reagent afforded a benzyl alcohol that underwent "ionic hydrogenation" by stirring with triethylsilane in the presence of trifluoroacetic acid (27). Deprotection of the methyl ether afforded 5-alkyl derivatives. This Grignard method was also harnessed in the synthesis of derivatives

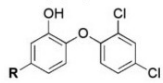
TABLE 1
Inhibitory properties of selected 5-substituted derivatives



Compound	R	EC_{50} [a] 3D7/Dd2 (μM)	PfENR [a] IC_{50} (nM)
triclosan	Cl	2.8 ± 0.2 / 3.8 ± 0.2	73 ± 21
2	CH_3	10 ± 1 / 14 ± 3	200 ± 60
3	Ph	2.3 ± 0.3 / 3.9 ± 1.5	140 ± 40
4	CN	19 ± 1 / 22 ± 4	49 ± 20
5	2H-tetrazol-5-yl	>100 / >100	>100,000
6	$C(O)NH_2$	40 ± 9 / 45 ± 17	21,000 ± 9,000
7	COOH	>100 / >100	>100,000

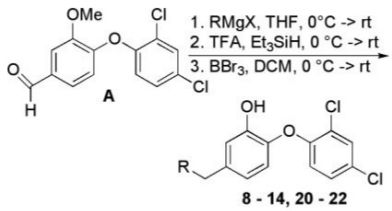
[a] Values are reported as mean ± S.E.

TABLE 2
Inhibitory properties of selected 5-alkyl derivatives



Compound	R	EC_{50} [a] 3D7/Dd2 (μM)	PfENR [a] IC_{50} (nM)
2	CH_3	10 ± 1 / 14 ± 3	200 ± 60
8	CH_2CH_3	4.6 ± 0.4 / 6.0 ± 1.1	110 ± 30
9	$(CH_2)_2CH_3$	2.5 ± 0.2 / 3.0 ± 0.5	210 ± 60
10	$(CH_2)_3CH_3$	3.5 ± 0.3 / 3.2 ± 1.0	480 ± 80
11	$CH_2CH(CH_3)_2$	2.0 ± 0.1 / 2.8 ± 0.6	180 ± 80
12	$CH_2CH(CH_3)CH_2CH_3$	2.0 ± 0.2 / 2.9 ± 0.6	290 ± 20
13	$(CH_2)_2CH(CH_3)_2$	7.4 ± 1.8 / 8.0 ± 0.6	120 ± 50
14	cyclohexylmethyl	3.4 ± 0.3 / 4.8 ± 0.6	530 ± 100

[a] Values are reported as mean ± S.E.

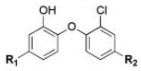


SCHEME 1. **5-Substituted alkyl derivatives.**

20–**22** in Table 3, whereas use of the appropriate lithiated pyridines (28) afforded **26**–**28**. Synthesis of the aryl and heteroaryl analogs of **3** shown in Table 3 relied on the transformation of 4-bromo-2-methoxyphenol or 5-bromotriclosan via variants of the Suzuki-Miyaura reaction (26, 29, 30).

Activation of Isoniazid (INH) and Synthesis of INH-NAD—Purified InhA (2 μM) was incubated with INH (500 μM), NAD^+ (500 μM), and Mn(III) pyrophosphate (500 μM) in 2 ml of sodium phosphate buffer for 1 h. After InhA inhibition, the enzyme solution was concentrated to 0.3 mM and purified by a Superdex™ 200 column preequilibrated with 50 mM phos-

TABLE 3
Inhibitory properties of selected 5-aryl and heteroaryl derivatives



Compound	R_1	R_2	EC_{50} [a] 3D7/Dd2 (μM)	PfENR [a] IC_{50} (nM)
3	Ph	Cl	2.3 ± 0.3 / 3.9 ± 1.5	140 ± 40
15	*o*-CH_3-Ph	Cl	1.6 ± 0.3 / 2.5 ± 0.6	440 ± 100
16	*o*-CH_3-Ph	CN	2.6 ± 0.3 / 3.3 ± 0.6	410 ± 100
17	*m*-CH_3-Ph	Cl	3.5 ± 0.3 / 6.1 ± 1.4	230 ± 70
18	*p*-CH_3-Ph	Cl	2.9 ± 0.6 / 4.3 ± 1.1	190 ± 70
19	*p*-F-Ph	Cl	2.1 ± 0.5 / 3.1 ± 1.3	38 ± 10
20	CH_2Ph	Cl	11 ± 2 / 12 ± 2	71 ± 10
21	$(CH_2)_2$Ph	Cl	5.6 ± 2.0 / 11 ± 1	76 ± 10
22	$(CH_2)_3$Ph	Cl	6.4 ± 2.1 / 7.2 ± 2.7	440 ± 200
23	2-pyridyl	CN	8.1 / 12 [b]	700 ± 50
24	3-pyridyl	Cl	23 ± 7 / 39 ± 12	33,000 ± 5,000
25	4-pyridyl	CN	11 ± 2 / 11 ± 2	2,800 ± 1,000
26	CH_2(2-pyridyl)	Cl	7.5 ± 0.6 / 13 ± 1	640 ± 100
27	CH_2(3-pyridyl)	Cl	6.9 ± 0.3 / 11 ± 3	840 ± 300
28	CH_2(4-pyridyl)	CN	7.4 ± 0.3 / 13 ± 2	530 ± 100

[a] Values are reported as mean ± S.E.
[b] $n = 1$.

phate buffer (pH 7.5). InhA in complex with inhibitor was then concentrated and heated for 40 s at 100 °C. After the heat treatment, the adduct was separated from denatured enzyme by filtration (Amicon cut-off size 30 kDa). The concentration of the inhibitor was determined by its absorbances at 260 and 326 nm, according to the procedure described by Nguyen (31).

PfENR Enzymatic Inhibition Assay—PfENR inhibition assays were carried out on a Cary 100 Bio Spectrophotometer or POLARstar Optima at 25 °C by monitoring oxidation of NADH at 340 nm as described previously (10, 15). Briefly, reactions were initiated by adding crotonoyl-CoA (300 μM) to assay mixtures containing buffer (150 mM NaCl and 20 mM Tris at pH 7.9), PfENR (50 nM), NADH (400 μM), NAD^+ (40 μM), and inhibitor that had been preincubated for 30 min. The IC_{50} was determined from a dose-response plot of enzyme fractional activity as a function of inhibitor concentration.

PfENR Enzymatic Inhibition Time Course Studies—These were carried out on a Cary 100 Bio Spectrophotometer by monitoring oxidation of NADH at 340 nm for 5 min after initiation of the enzymatic reaction. Two different reaction types were profiled. The no-incubation type of reaction was initiated by adding 50 nM PfENR to buffer (150 mM NaCl and 20 mM Tris at pH 7.9) and then adding a mixture of 40 μM NAD^+, 400 μM NADH, 300 μM crotonoyl-CoA, and inhibitor (respective concentrations are shown below). The preincubation variant of this reaction was conducted by adding 50 nM PfENR to buffer (150 mM NaCl and 20 mM Tris at pH 7.9), 40 μM NAD^+, 400 μM NADH, and inhibitor at concentrations listed below. This mixture was held for 30 min, and the enzymatic reaction was initiated by adding 300 μM crotonoyl-CoA. The time course plots were then constructed by graphing [NAD^+] *versus* time. The [NAD^+] at each time point was calculated as follows: $[NAD^+] = 0.040 + (0.40 - (A_{340}/6.3))$, where A_{340} is the measured absorbance at 340 nm and 6.3 cm^{-1} mM^{-1} is the NADH extinction coefficient. These plots are shown in the supplemental data. Inhibitor concentrations were as follows: [**2**] = 200 nM, [**4**] = 94 nM, [**15**] = 440 nM, [**19**] = 54 nM, [**20**] = 78 nM, [**21**] = 87 nM, [**22**] = 660 nM, [**23**] = 1.5 μM, [**26**] = 640 nM, and [**27**] = 530 nM.

Crystallization Details for PfENR·INH-NAD—Purified PfENR (2 μM) was incubated with INH (500 μM), NAD^+ (500 μM), and Mn(III) pyrophosphate (500 μM) in 2 ml of sodium phosphate buffer for 1 h. After PfENR inhibition, the enzyme solution was concentrated to 0.3 mM and used for crystallization. PfENR·INH-NAD was crystallized in hanging droplets containing 2 μl of protein solution at 10 mg/ml and 2 μl of buffer (2.4 M $(NH_4)_2SO_4$, 0.1 M MES, pH 5.6) at 16 °C in Linbro plates against 1 ml of the same buffer. Protein crystals formed after ~4 days with a space group of $P4_32_12$. Data were collected at 121 K using paratone as cryoprotectant. Crystals of PfENR·INH-NAD diffracted X-rays to 2.5 Å with a Bruker Proteum CCD Detector system (Proteum 6000) coupled with a Rigaku generator (MM007 HF). The data were integrated and reduced using Proteum software. Crystals produced from PfENR in complex with INH-NAD were isomorphous to those of the native enzyme. Initial phases were obtained by molecular replacement using the apo-PfENR structure (Protein Data Bank 1VRW) and refined with CNS (32). $F_o - F_c$ and $2F_o - F_c$ electron density maps were calculated, and additional density resembling the INH-NAD was found. The ligand was incorporated into the additional density, and the whole model was rebuilt using XtalView (33). During the final cycles of the refinement, water molecules were added into peaks above 3-σ of the $F_o - F_c$ electron density maps that were within hydrogen-bonding distances from the appropriate protein atoms. The final model was refined to an R_{factor} of 21% and an R_{free} of 27%. At the end of the refinement and model building, a final simulated annealing was conducted to eliminate overrefinement, for this and all other models reported in this study.

*Crystallization Details for **15***—Since crystals of the PfENR·NAD^+·**15** complex were isomorphous to those of the native enzyme, the protein coordinates of the PfENR·NADH complex (Protein Data Bank code 1VRW) were used to calculate $2F_o - F_c$ and $F_o - F_c$ maps. The $F_o - F_c$ map calculated at 2.5 Å resolution identified additional density consistent with the structure of **15**. The inhibitor was built into the additional density with XtalView (33), and subsequent refinement with REFMAC (34) gave rise to an R_{work} of 20% and an R_{free} of 26% after placement of ordered water molecules.

*Crystallization Details for **21–23** and **26***—The crystallization conditions utilized the hanging drop and vapor diffusion methods. Typically, 24 mg/ml PfENR was incubated with 4 mM NADH in 20 mM Tris, pH 8.0, 150 mM NaCl, 10 mM EDTA, and 1 mM dithiothreitol. 2 μl of this mixture was mixed with 2 μl of the well solution consisting of 2.5 M $(NH_4)_2SO_4$, 100 mM NaOAc, and 200 mM Tris, pH 7.7, and equilibrated at 17 °C to afford the binary complex. The triclosan analog was dissolved in acetonitrile and directly added to this crystallization drop containing crystals of binary complex with a space group $P4_32_12$ to a final concentration of 2.1 mM. Diffraction data were collected after 10 days of soaking at 120 K using a Raxis IV++

imaging system outfitted with double focusing mirrors. The data were collected and processed with CrystalClear/*d*TREK* (35) and HKL2000 (36) software and refined with CCP4 (37) and Phenix (38). For all of the crystallization studies, the data collection and refinement statistics may be found in the supplemental data.

In Vitro Drug Assays with Cultured P. falciparum—Compounds were tested for antimalarial activity against cultured *P. falciparum* parasites (Dd2 and 3D7 strains) using a [^{3}H]hypoxanthine uptake method (39). This measures [^{3}H]hypoxanthine incorporation as a function of drug concentration and is used to derive EC_{50} values, representing compound concentrations that inhibit parasite growth by 50%. Assays were performed in duplicate with synchronized ring stage cultures for 72 h on at least three separate occasions, with the exception of **23**, which was tested once. [^{3}H]Hypoxanthine was added for the last 24 h (40). Culture media consisted of RPMI 1640 with L-glutamine (Invitrogen), 50 mg/liter hypoxanthine, 10 mg/liter gentamycin, 25 mM HEPES, 0.225% $NaHCO_3$, and 0.5% Albumax I (Invitrogen), with Albumax providing a source of exogenous fatty acids.

Expression of katG in P. falciparum—This gene was expressed using the Bxb1 integrase system, recently developed to enable rapid, site-specific transgene integration in *P. falciparum* (41). Here, the transgene expression plasmid pLN-ENR-V5, which is identical to pLN-ENR-GFP (41) except that the GFP sequence was replaced by a V5 epitope tag, was adapted to drive *katG* expression from the ring stage *hrp3* promoter. This 1.9-kb promoter was amplified using the primers 5′-taaGCATGCCCCAGTCACGACGTTGTAAAACG (SphI site underlined) and 5′-taaCCTAGGTTTTTTAAAATAACTGTATTAATATAAAA (AvrII site underlined), cloned into pGEM-T (Promega), and then subcloned as an SphI-AvrII fragment into pLN-ENR-V5 digested with SphI and AvrII to remove the calmodulin promoter. *katG* was amplified from *M. tuberculosis* genomic DNA using the primers 5′-taaCCTAGGATGTCCGAGCAACACCCACCC (AvrII site underlined) and 5′-taaCGTACGTCAGCGCACGTCGAACCTGTC (BsiWI site underlined) and cloned into pGEM-T. An error-free insert was identified following double-stranded sequencing and subcloned as a 2.2-kb AvrII-BsiWI fragment into the *hrp3* construct digested with AvrII and BsiWI to remove the *pfenr* gene without disrupting the V5 tag. The resulting 9.5-kb vector, denoted pLN-hrp3-katG-V5, was electroporated into Dd2_attB parasites along with the pINT plasmid that expresses Bxb1 integrase. Integrase-mediated insertion of pLN-hrp3-katG-V5 into the chromosomal attB site was demonstrated by PCR as described (41). Western blot analysis with antibodies to the V5 epitope tag was performed with parasite extracts prepared from the *katG*-expressing line or the parental control Dd2_attB line, as described (41).

RESULTS AND DISCUSSION

Our design efforts at the triclosan 5-position were initially aided by concurrent investigations in our laboratories to study the inhibition of PfENR by INH (Fig. 1). INH, a front-line antitubercular drug used since 1952 (42, 43), inhibits InhA via the formation of a covalent adduct between activated INH and

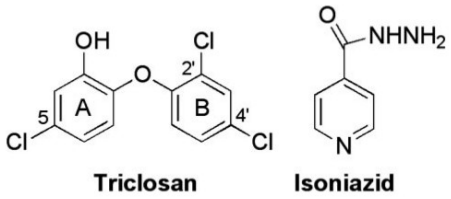


FIGURE 1. **Chemical structures of triclosan and isoniazid.**

NADH (44–47). Consistent with the 30% sequence identity between InhA and PfENR (10), the INH-NAD adduct was determined to inhibit the PfENR enzyme with an IC_{50} of 320 ± 40 nM. INH was reported to have modest activity against *P. falciparum* parasites cultured *in vitro* (48) and, in our hands, exhibited EC_{50} values of 1.4 ± 0.2 and 1.9 ± 0.1 mM *versus* the *P. falciparum* 3D7 and Dd2 strains, respectively.

Given the requirement for activation of the INH prodrug in *M. tuberculosis* by the catalase peroxidase KatG (49), we tested whether expression of *katG* in *P. falciparum* would result in activation of the INH prodrug and an increase in potency against the cultured malarial parasites. Using site-specific integration techniques (41), recombinant parasites were obtained in which *M. tuberculosis katG* was integrated into the genome, as confirmed by PCR and Southern blot analyses. Expression of *katG* was also confirmed by Western blot using a monoclonal antibody to a V5 epitope tag that had been appended to the C-terminal end of KatG. This study was designed to express KatG in the relatively voluminous parasite cytoplasm, where it could activate INH prior to the drug attaining the apicoplast, and did not include a parallel experiment to direct KatG to the apicoplast using an apicoplast targeting sequence (3). Drug assays with INH yielded IC_{50} values of 1.8–2.0 mM for Dd2 parasites expressing *katG* as well as for the control lines Dd2 and Dd2_attB (41). A similar result was obtained for lines created with the 3D7 strain, with IC_{50} values of 1.2–1.7 mM for the *katG*-expressing and control lines. The very weak antiparasitic activity of INH is inconsistent with its significant inhibition of purified PfENR and may be due to a lack of proper activation of INH within the malarial parasite or possibly because INH activity against cultured parasites is not caused by inhibition of PfENR.

Despite the lack of whole-cell activity, we were able to generate the INH-NAD adduct bound to PfENR. PfENR was crystallized in the presence of NAD^+, and INH was activated through oxidation by manganese(III) pyrophosphate (50). The overall structure (Fig. 2*A*), including the nucleotide binding site and the proposed fatty acyl substrate binding pocket of PfENR·INH-NAD, is similar to that observed in the structure of PfENR·NADH (10). The INH-NAD binding region includes the same nucleotide binding site observed in PfENR·NADH and an extended small pocket housing the isonicotinoyl moiety. This small pocket, partially overlapping with the purported fatty acyl substrate-binding site, is lined predominantly by side chains of hydrophobic residues Tyr-267, Tyr-277, Gly-313, Pro-314, Phe-368, Ile-369, Ala-372, and Ile-373. This pocket also serves as a small portal to the solvent at one side of the active site,

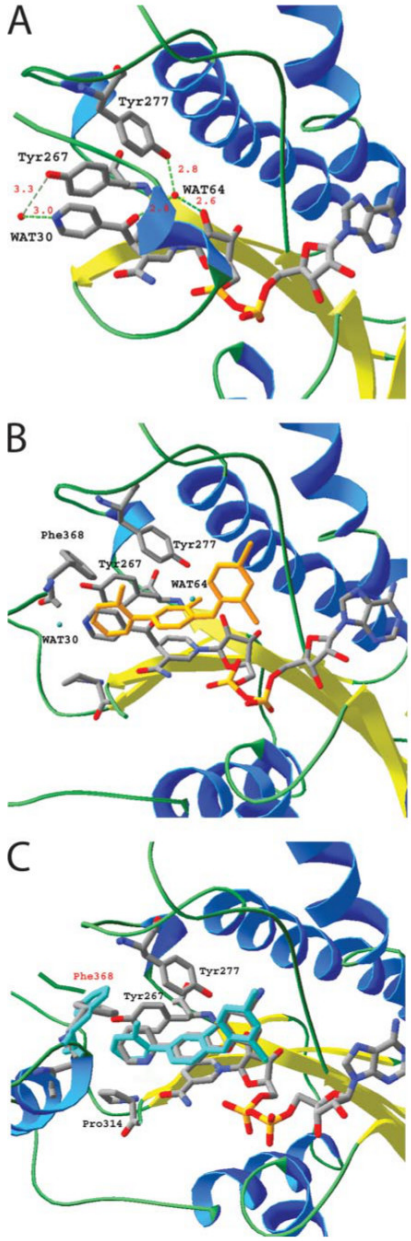


FIGURE 2. **X-ray crystal structures of ligands bound to PfENR.** *A*, INH-NAD (*stick drawing*) bound to PfENR (*ribbon* and *tube*). The INH isonicotinamide pyridyl nitrogen and the Tyr-267 side chain phenol are hydrogen-bonded to a water molecule (WAT30). The side chain phenol of Tyr-277, the carbonyl of the INH isonicotinamide moiety, and NADH ribose hydroxyl are all hydrogen-bonding to a water molecule (WAT64). Distances shown are in angstroms. *B*, **15** (*stick drawing*; *gold coloring*) bound to PfENR (*ribbon* and *tube*) in the presence of NAD^+ (*space fill*), with an overlay of INH-NAD (*stick drawing*) bound to PfENR. The enzyme active site accommodates the 5-substituent most evidently by rotation of Phe-368 and movement of Tyr-267. *C*, **23** (*stick drawing*) bound to PfENR (*ribbon* and *tube*) and NAD^+ (*stick drawing*), with an overlay of **15** (*stick drawing*; *turquoise blue coloring*) bound to PfENR. Phe-368 is *colored* according to the ligand bound to PfENR·NAD^+ atom *color* (**23** bound) or *turquoise blue* (**15** bound). The 2-pyridyl moiety of **23** is engaged in hydrophobic interactions with Pro-314, Phe-368, Ile-369, and Ala-372 and a face-to-face π-stacking interaction with Tyr-267. For all PfENR structures in Figs. 2 and 3, key residues are noted in *stick format*, and parts of the structure are removed for clarity. Both figures were made using the program SwissPDB viewer (63).

where the isonicotinoyl group of the adduct points toward the entrance. Comparison of the structures of PfENR·NADH and PfENR·INH-NAD reveals that the side chain of Tyr-267 has rotated ~15° to form an aromatic ring-stacking interaction with the pyridine ring of the isonicotinoyl group. The isonicotinoyl nitrogen, oxygen of the Tyr-267 hydroxyl group, and an ordered water molecule interact in a triangle hydrogen-bonding network. A similar conformational change is observed with Tyr-277, whose side chain has also twisted ~15° to form a hydrogen bond with a highly conserved water molecule located 2.8 Å away. This water molecule is in the center of a triangle hydrogen-bonding network formed by the isonicotinoyl carbonyl oxygen, 2′-hydroxyl oxygen of the co-factor ribose, and oxygen of the Tyr-277 hydroxyl group.

Superimposition of the PfENR·INH-NAD and PfENR·triclosan structures revealed that this ordered water molecule is in the same location as the hydroxyl group of triclosan. Although the structure of PfENR bound with its substrate has yet to be defined, it is very likely that this water occupies the binding site of the carbonyl group of the fatty acyl substrate, based on analysis of the complex structure of InhA bound with fatty acyl substrate (51). Although the isonicotinoyl group and the A-ring of triclosan both fit inside the hydrophobic pocket, the former extends ~3 Å deeper into the pocket along a vector most easily reached from the 5-position of triclosan. The introduction of a bulky 5-substituent could block the binding of the fatty acyl substrate, fill the enzyme hydrophobic pocket, and displace the bound water molecule. These observations support our hypothesis that hydrophobic 5-position substituents may be highly desirable.

To test our 5-position hypothesis further, a range of differentially substituted 5-position analogs was synthesized and examined for activity *versus* PfENR and cultured parasites (Table 1). All final compounds (except for **23**) were tested on at least three separate occasions to determine their inhibition of purified PfENR enzymatic activity and parasite whole-cell growth via a [^{3}H]hypoxanthine uptake assay (39, 40).

A clear preference for hydrophobic groups was established at the outset. Methyl **2** is slightly less potent than triclosan against purified enzyme, whereas it is ~4-fold less potent *versus* the cultured parasites. Phenyl analog **3**, although marginally less active against the enzyme, is essentially equipotent to triclosan in the parasite assay. In contrast, analogs with polar functionality, such as tetrazole (**5**), carboxamide (**6**), and carboxylic acid (**7**), are much less efficacious in both assay systems than tri-

closan. Carboxylic acid **7** was prepared and tested by Chhibber *et al.* (16) and demonstrated a similar lack of antiparasitic activity. Intriguingly, in our hands, **7** is inactive against PfENR, having an IC_{50} of >100 μM ($n = 4$ independent trials) compared with their reported value of 560 nM. The potent enzyme inhibition of **4** is consistent with our design hypothesis, since the 5-nitrile moiety should extend deeper into the proximal binding pocket than the triclosan 5-chloro group, with molecular modeling suggesting the potential for an energetically favorable hydrogen-bonding interaction between the nitrile nitrogen and the hydroxyl group of Tyr-267. In addition, **4** was estimated to be only slightly less hydrophobic (<1 log unit) than triclosan, based on AlogP and MlogP values (calculated using Accord for Excel, Version 6.1, Accelrys Software Inc.; data not shown). The reduced hydrophobicity may explain why the antiparasitic activity of **4** is diminished compared with that of triclosan. For our compound series (Tables 1–3), a modest correlation ($r^2 = 0.55$) was observed between antiparasitic activity and the calculated hydrophobicity of triclosan and its 5-substituted analogs, as predicted by MlogP and AlogP values.

Overall, the results in Table 1 are consistent with our observations concerning the 5-position binding pocket derived from the x-ray crystal structure of triclosan bound to the PfENR·NAD^+ complex. Analogs **2** and **3** represented reasonable origins from which to test our hypothesis that suitable hydrophobic groups should better fill the 5-position pocket than the triclosan 5-chloro group.

While attempts were undertaken to determine the x-ray structures of bound **2** and/or **3** to aid design efforts, we prepared a number of alkyl-substituted triclosan derivatives (Table 2) to expand the SAR at the 5-position. Compounds **8**–**14** all achieve an improvement in antiparasitic activity over methyl **2**. It is interesting to note that the introduction of a cyclic constraint (**11**/**12** → **14**) does not consistently improve potency in the parasite and enzyme assays. Although **8** and **13,** in addition, improve on the enzymatic potency of methyl **2,** they are slightly less efficacious than triclosan. These 5-alkyl derivatives do not demonstrate a correlation between enzyme inhibition and antiparasitic efficacy. This may be due to a number of factors, including differences in their pharmacokinetic profiles and potential off-target activities.

The initial examination of analogs of 5-phenyl **3** began with the tolyl compounds **15**–**18** (Table 3). The *o*-tolyl analog **15** is the most potent 5-substituted analog *versus* cultured parasites prepared to date, being 2-fold more active than triclosan. In contrast, the placement of a methyl group on the phenyl ring of **3** has a detrimental effect on enzyme activity. This structural perturbation affects parasite growth inhibition in a manner dependent on the methyl group location. It is interesting to note that *p*-F-phenyl derivative **19** is the most potent triclosan analog we have prepared to date *versus* purified enzyme, being 2-fold more active than triclosan. **19**, also slightly more active than triclosan against both strains of parasite, is proposed through molecular modeling studies to attain its enhanced binding affinity to PfENR via a hydrogen bond between the fluoro group and the proximal hydroxyl moiety of Tyr-267.

Comparison of the structure of **15** (Fig. 2*B*) with that of bound triclosan clearly shows that the 5-*o*-tolyl moiety better fills the enzyme hydrophobic pocket into which the triclosan 5-chloro group projects. This results in an increase in surface area of interaction of ~50 $Å^2$ between **15** and the enzyme active site, compared with the binding of triclosan. The *o*-tolyl group has enhanced van der Waals interactions with Pro-314 and Phe-368. Phe-368 has rotated ~60° from its conformation in the triclosan structure to engage the *o*-tolyl moiety in an edge-to-face interaction. Tyr-267 is now participating in a face-to-face interaction ($d_{\text{centroid-centroid}} = 3.7$ Å) with the *o*-tolyl group and has been displaced from C6 of the triclosan A-ring by ~0.5 Å. Due to its *o*-tolyl moiety, **15** is also engaged in van der Waals interactions with Ile-323 of the putative substrate-binding loop and Ala-372. The chloro group of triclosan is incapable of making these interactions. Comparison of the x-ray structure of bound **15** with that of PfENR·INH-NAD (Fig. 1*A*) is aided by an overlay of the two structures. The two ligands clearly both occupy the hydrophobic binding pocket while extending into the proposed substrate-binding pocket.

Despite the presence of these novel and presumably energetically favorable contacts between **15** and PfENR, we note that **15** is six times less potent than triclosan against purified enzyme. The primary interaction of triclosan with PfENR is that of the phenol with the 2′-hydroxyl of the ribose unit and the hydroxyl of Tyr-277 (10). The metrics describing these interactions, and those pertinent to the face stacking of the cofactor nicotinamide ring with the A-ring of the diaryl ether scaffold, are essentially the same for **15** and triclosan. It is plausible that the binding of **15** to the PfENR·NAD^+ complex results in unfavorable interactions not observed in the analysis of the x-ray structure.

In an attempt to better mimic the INH pyridyl moiety, analogs **23**–**25** were prepared and assayed. It is interesting to note that replacement of the *o*-tolyl group in **15** with a 2-, 3-, or 4-pyridyl increases neither enzyme binding affinity nor antiparasitic activity. These compounds are also significantly less active than INH *versus* PfENR. Noticeably, 3-pyridyl analog **24** is the least active of the trio.

The complex structure of PfENR·NAD^+·**23** (Fig. 2*C*) readily aligns with that of the corresponding *o*-tolyl analog **15**. It should be noted that the 4′-substituent (Cl *versus* CN) does not significantly perturb the overall structures. The A-ring and B-ring atoms of the triclosan scaffolds nearly overlay, whereas the conformations of the 5-substituents are clearly different. The 2-pyridyl moiety of **23** has flipped 40° from the position of the *o*-tolyl group of **15**. Correspondingly, Phe-368 returns close to its position in the structure with bound triclosan by rotating 45° from its position in the *o*-tolyl structure, obviating the possibility of an edge-to-face interaction with the 5-substituent. The 2-pyridyl group still maintains van der Waals contact with Phe-368 in addition to Pro-314, Ile-369, and Ala-372. Similar to the structure of bound **15**, Tyr-267 has moved 0.4 Å away from C6 of the triclosan A-ring to facilitate a face-to-face interaction with the 5-(2-pyridyl) moiety.

A combination of two prior strategies was then explored: use of an alkyl linker in conjunction with an aryl or pyridyl group. It was hoped that the PfENR active site would exhibit further flexibility and that enhanced interactions with the 5-substituent would confer considerable improvements in compound

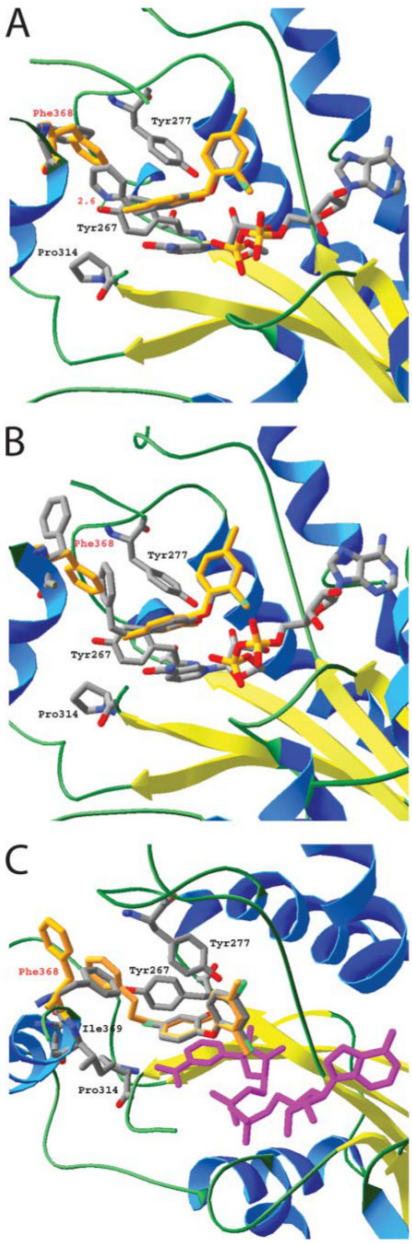


FIGURE 3. **X-ray crystal structures of ligands bound to PfENR.** *A*, **26** (*stick drawing*) bound to PfENR (*ribbon* and *tube*) and NAD^+ (*stick drawing*), with an overlay of triclosan (*stick drawing*; *gold coloring*) bound to PfENR. Phe-368 is *colored* according to the ligand bound to PfENR·NAD^+ atom *color* (**26** bound) or *gold* (triclosan bound). The 2-pyridyl moiety of **26** is engaged in hydrophobic interactions with Pro-314, Ile-323, Phe-368, Ile-369, and Ala-372 and a hydrogen bond ($d_{N\text{-}O}$ = 2.6 Å) with the hydroxyl moiety of Tyr-267. *B*, **21** (*stick drawing*) bound to PfENR (*ribbon* and *tube*) and NAD^+ (*stick drawing*), with an overlay of triclosan (*stick drawing*; *gold coloring*) bound to PfENR. Phe-368 is *colored* according to the ligand bound to PfENR·NAD^+ atom *color* (**26** bound) and *gold* (triclosan bound). The 5-phenethyl moiety interacts via van der Waals contacts with Val-274, Tyr-277, Pro-314, Ala-320, Ile-323, Phe-368, Ile-369, and Ala-372. *C*, **22** (*stick drawing*; *gold coloring*) bound to PfENR (*ribbon* and *tube*) and NAD^+ (stick drawing; *magenta coloring*), with an overlay of triclosan (*stick drawing*) bound to PfENR. Phe-368 is *colored* according to the ligand bound to PfENR·NAD^+ atom *color* (triclosan bound) or *gold* (**22** bound). The 5-phenylpropyl moiety interacts via van der Waals contacts with Val-274, Gly-276, Tyr-277, Pro-314, Ala-320, Ile-323, Phe-368, Ile-369, and Ala-372.

potency. Examination of analogs of the type $(CH_2)_nPh$ (n = 1–3) demonstrated that benzyl **20** and phenethyl **21** are essentially equipotent to triclosan against PfENR. This increase in enzyme inhibition upon one carbon homologation is also observed in the pyridyl series. Interestingly, further homologation of **21** to **22** affords a much weaker PfENR inhibitor while not significantly altering activity against cultured parasites.

Analysis of the x-ray crystal structure of **26** (Fig. 3*A*) demonstrates how the enzyme active site similarly accommodates a group at the 5-position larger than chloride. Phe-368 has moved in deference to the 2-pyridylmethyl group, interacting instead with the 5-substituent aromatic ring in an edge-to-face mode. By virtue of its formal one-carbon homologation from **23**, **26** is able to utilize the pyridyl group nitrogen to form a hydrogen bond with the hydroxyl group of Tyr-267 instead of participating in a π-stacking interaction. The 5-pyridylmethyl group additionally participates in hydrophobic interactions with Pro-314, and the purported substrate binding loop residue Ile-323, Ile-369, and Ala-372.

The bound structure of **21** (Fig. 3*B*) is instructive in demonstrating the ability of the active site to house the even larger 5-phenethyl moiety. The A- and B-ring skeletons of **21** and triclosan overlap well, whereas the 5-phenethyl moiety extends deep into the proximal hydrophobic pocket lined by Val-274, Tyr-277, Pro-314, Ala-320, Ile-323 of the proposed substrate binding loop, Phe-368, Ile-369, and Ala-372. Phe-368 rotates more than 90° from its position in the respective triclosan structure such that only its side chain C_α is interacting with the 5-position pendant phenyl group.

One-carbon homologation of the alkyl tether in **21** affords phenylpropyl **22**, where the x-ray structure (Fig. 3*C*) demonstrates further flexibility in the PfENR active site. The structure is quite similar to that of bound **21,** with the main deviation occurring due to extension of the phenylpropyl side chain further into the hydrophobic pocket. The result is an additional displacement of Phe-368 by 0.5 Å to accommodate the larger 5-substituent and an interaction with Gly-276.

Overall, these x-ray structures allow a detailed examination of the structural effects of increasing the size of the triclosan 5-substituent from that of a chloride to an aryl/pyridyl group spaced between zero and three carbons from the A-ring. This comparison illustrates that the most obvious way the enzyme accommodates the larger 5-substituent is by movement of the Phe-368 side chain. The effect on Phe-368 becomes more pro-

nounced as the size of the 5-substituent is increased. We do not, however, observe significant changes in the key metrics describing the π-stacking of the A-ring with the nicotinamide ring and the hydrogen bonding of the phenol with the ribose 2′-hydroxyl and Tyr-277 hydroxyl.

Finally, having noted the structural perturbation of the 5-substituent on the PfENR active site, we wished to determine if a correlation exists between the nature of this moiety and its compound's binding kinetics. The time course of the PfENR reduction of crotonoyl-CoA was followed spectrophotometrically in the presence and absence of inhibitor (see "Experimental Procedures" and supplemental data). A subset of compounds (IC_{50} values *versus* PfENR of 38–840 nM) was selected with a variety of 5-substituents: cyano **4**, alkyl (methyl **2**), aryl (*o*-tolyl **15**; *p*-F-phenyl **19**; benzyl **20**; phenethyl **21**; phenylpropyl **22**), and heteroaryl (2-pyridyl **23**; 2-pyridylmethyl **26**; 3-pyridylmethyl **27**). The compounds demonstrate what may be described as a range of behaviors varying between slow binding and classical reversible kinetics (52). 2-Pyridyl **23** and 2-pyridylmethyl **26** most closely exemplify classical reversible inhibition. It should be noted that careful inspection of the respective x-ray crystal structures of **23** and **26** bound to the PfENR·NAD^+ complex did not uncover significant differences as compared with the previously reported structures of PfENR·NAD^+·triclosan and PfENR·NADH (10). Methyl **2**, *o*-tolyl **15**, benzyl **20**, phenylpropyl **22**, and 3-pyridylmethyl **27,** on the contrary, clearly exhibit slow binding kinetics. With these slow binders, the preincubation of enzyme with compound, NAD^+, and NADH for 30 min led to a significant decrease in catalytic activity as compared with the use of no preincubation period. Triclosan also has been reported to behave similarly (8, 53, 54). Clearly, the placement of a 5-substituent on triclosan has an effect on the binding kinetics with respect to PfENR that, at this moment, does not appear to follow a trend that is structure- or activity-based.

In this study, a series of 5-substituted triclosan derivatives has been prepared and assayed for inhibition of PfENR and parasite growth. The efforts described herein significantly expand the knowledge of triclosan analog SAR at the 5-position while resulting in 2-fold improvements in the antiparasitic and/or enzyme potencies of triclosan. Further profiling of these compounds will require an analysis of whether they also inhibit mammalian enzymes, especially in light of recent reports documenting the existence of FAS-II pathway in the mitochondria in mammalian cells (19–22). We note that changes in enzyme and parasite whole-cell inhibition data do not always correlate, suggesting that properties other than inhibition of PfENR under cell-free *in vitro* conditions can influence potency in whole-cell assays. These could include differences between compounds in their levels of drug accumulation and access to target or off-target effects in cultured parasites.

A comparison of our whole-cell and enzyme inhibition, performed for inhibitors whose enzyme IC_{50} values were submicromolar, revealed that IC_{50} values were 4–400-fold lower than whole-cell EC_{50} values (with the geometric mean of the -fold difference being 20). Similar discrepancies, attributed to inefficient uptake of compounds into the parasite, have been reported with inhibitors of the *P. falciparum* cysteine proteases (55) or dihydroorotate dehydrogenase (56). Limited cell permeability could equally be a cause for the discrepancy between whole-cell and enzyme inhibition values observed with triclosan and its analogs, particularly in view of the requirement for apicoplast inhibitors to traverse four membranes surrounding this organelle (4, 24) in addition to the host cell and parasite membranes.

The 5-substituted triclosan derivatives presented herein offer an expansion of the general diarylether phenol chemotype as previously explored by other researchers and ourselves (10, 15–17). It is interesting to compare these triclosan analogs *versus* other known inhibitors of PfENR. Most closely related to the triclosan series is a single less potent diphenylamine, prepared in a directed attempt to replace the diaryl ether oxygen of triclosan (10). A high throughput screen of small molecules for activity *versus* InhA uncovered an amide-substituted indole and a disubstituted pyrazole that also afforded modest inhibition of purified PfENR and cultured parasites (11). In the case of the amide-substituted indole, the central amide moiety is most probably a surrogate for the phenol group of triclosan. An independent study subsequently reported modest PfENR inhibition with a remarkably similar set of pyrazoles (57). These researchers also reported the discovery of a rhodanine class of small molecule binders to PfENR, with the most efficacious member of this class approximating the activity of triclosan (58). Building on initial investigations into plant polyphenols as antibacterial FAS-II inhibitors (59), other researchers have independently examined natural product inhibitors of PfENR where a common structural theme is the presence of one or more phenolic moieties (60, 61). Interestingly, select members of the green tea catechin family were high nanomolar inhibitors of PfENR alone while also potentiating the efficacy of triclosan inhibition proposedly through the formation of a PfENR·catechin·triclosan complex (62).

Our experimental results stand in sharp contrast to a recent proposal that a 5-chloro group is optimal for enzyme inhibition due to its small size (16). Against PfENR, 5-*p*-fluorophenyl **19** is 2 times more potent than triclosan, and the corresponding 5-benzyl **20** and 5-phenethyl **21** analogs are equipotent to triclosan. These groups are all larger than chloride, and x-ray crystallographic studies demonstrate that the enzyme can undergo conformational changes to support moieties larger than a chloride. Future efforts will seek to utilize the insights from this structural study to design more potent inhibitors of PfENR and parasite growth. This will be done in conjunction with probing other positions along the diaryl ether scaffold of triclosan.

REFERENCES

1. Snow, R. W., Guerra, C. A., Noor, A. M., Myint, H. Y., and Hay, S. I. (2005) *Nature* **434,** 214–217
2. Greenwood, B., and Mutabingwa, T. (2002) *Nature* **415,** 670–672
3. Waller, R. F., Keeling, P. J., Donald, R. G. K., Striepen, B., Handman, E., Lang-Unnasch, N., Cowman, A. F., Besra, G. S., Roos, D. S., and McFadden, G. I. (1998) *Proc. Natl. Acad. Sci.* **95,** 12352–12357
4. Ralph, S. A., van Dooren, G. G., Waller, R. F., Crawford, M. J., Fraunholz, M. J., Foth, B. J., Tonkin, C. J., Roos, D. S., and McFadden, G. I. (2004) *Nat. Rev. Microbiol.* **2,** 203–216
5. Lu, J. Z., Lee, P. J., Water, N. C., and Prigge, S. T. (2005) *Combinatorial Chem. High Throughput Screening* **8,** 15–26
6. McFadden, G. I., Reith, M. E., Munholland, J., and Lang-Unnasch, N.

(1996) *Nature* **381,** 482
7. Surolia, N., and Surolia, A. (2001) *Nat. Med.* **7,** 167–173
8. Kapoor, M., Reddy, C. C., Krishnasastry, M. V., Surolia, N., and Surolia, A. (2004) *Biochem. J.* **381,** 719–724
9. McLeod, R., Muench, S. P., Rafferty, J. B., Kyle, D. E., Mui, E. J., Kirisits, M. J., Mack, D. G., Roberts, C. W., Samuel, B. U., Lyons, R. E., Dorris, M., Milhous, W. K., and Rice, D. W. (2001) *Int. J. Parasitol.* **31,** 109–113
10. Perozzo, R., Kuo, M., Sidhu, A. S., Valiyaveettil, J. T., Bittman, R., Jacobs, W. R., Jr., Fidock, D. A., and Sacchettini, J. C. (2002) *J. Biol. Chem.* **277,** 13106–13114
11. Kuo, M. R., Morbidoni, H. R., Alland, D., Sneddon, S. F., Gourlie, B. B., Staveski, M. M., Leonard, M., Gregory, J. S., Janjigian, A. D., Yee, C., Musser, J. M., Kreiswirth, B., Iwamoto, H., Perozzo, R., Jacobs, W. R., Jr., Sacchettini, J. C., and Fidock, D. A. (2003) *J. Biol. Chem.* **278,** 20851–20859
12. Waller, R. F., Ralph, S. A., Reed, M. B., Su, V., Douglas, J. D., Minnikin, D. E., Cowman, A. F., Besra, G. S., and McFadden, G. I. (2003) *Antimicrob. Agents Chemother.* **47,** 297–301
13. Sharma, S., Sharma, S. K., Modak, R., Karmodiya, K., Surolia, N., and Surolia, A. (2007) *Antimicrob. Agents Chemother.* **51,** 2552–2558
14. Mi-ichi, F., Kita, K., and Mitamura, T. (2006) *Parasitology* **133,** 399–410
15. Freundlich, J. S., Anderson, J. W., Sarantakis, D., Shieh, H.-M., Yu, M., Valderramos, J.-C., Lucumi, E., Kuo, M., Jacobs, W. R., Jr., Schiehser, G. A., Fidock, D. A., Jacobus, D. P., and Sacchettini, J. C. (2005) *Bioorg. Med. Chem. Lett.* **15,** 5247–5252
16. Chhibber, M., Kumar, G., Parasuraman, P., Ramya, T. N. C., Surolia, N., and Surolia, A. (2006) *Bioorg. Med. Chem.* **14,** 8086–8098
17. Freundlich, J. S., Yu, M., Lucumi, E., Kuo, M., Tsai, H.-C., Valderramos, J.-C., Karagyozov, L., Jacobs, W. R., Jr., Schiehser, G. A., Fidock, D. A., Jacobus, D. P., and Sacchettini, J. C. (2006) *Bioorg. Med. Chem. Lett.* **16,** 2163–2169
18. Bhargava, H. N., and Leonard, P. A. (1996) *Am. J. Infect. Control* **24,** 209–218
19. Miinalainen, I. J., Chen, Z.-J., Torkko, J. M., Pirilä, P., Sormunen, R. T., Bergmann, U., Qin, Y.-M., and Hiltunen, J. K. (2003) *J. Biol. Chem.* **273,** 20154–20161
20. Zhang, L., Joshi, A. K., and Smith, S. (2003) *J. Biol. Chem.* **278,** 40067–40074
21. Zhang, L., Joshi, A. K., Hofmann, J., Schweizer, E., and Smith, S. (2005) *J. Biol. Chem.* **280,** 12422–12429
22. Witkowski, A., Joshi, A. K., and Smith, S. (2007) *J. Biol. Chem.* **282,** 14178–14185
23. Wang, L.-Q., Falany, C. N., and James, M. O. (2004) *Drug Metab. Dispos.* **32,** 1162–1169
24. Kohler, S., Delwiche, C. F., Denny, P. W., Tilney, L. G., Webster, O., Wilson, R. J. M., Palmer, J. D., and Roos, D. S. (1997) *Science* **275,** 1485–1489
25. Sullivan, T. J., Truglio, J. J., Boyne, M. E., Novichenok, P., Zhang, X., Stratton, C. F., Li, H.-J., Kaur, T., Amin, A., Johnson, F., Slayden, R. A., Kisker, C., and Tonge, P. J. (2006) *ACS Chem. Biol.* **1,** 43–53
26. Freundlich, J. S., and Landis, H. E. (2006) *Tetrahedron Lett.* **47,** 4275–4279
27. Kursanov, D. N., Parnes, Z. N., and Loim, N. M. (1974) *Synthesis* **9,** 633–651
28. Cai, D., Larsen, R. D., and Reider, P. J. (2002) *Tetrahedron Lett.* **43,** 4285–4287
29. Netherton, M. R., and Fu, G. C. (2001) *Org. Lett.* **3,** 4295–4298
30. Sindkhedkar, M. D., Mulla, H. R., Wurth, M. A., and Cammers-Goodwin, A. (2001) *Tetrahedron* **57,** 2991–2996
31. Nguyen, M., Quemard, A., Broussy, S., Bernadou, J., and Meunier, B. (2002) *Antimicrob. Agent Chemother.* **46,** 2137–2144
32. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **54,** 905–921
33. McRee, D. E. (1999) *J. Struct. Biol.* **125,** 156–165
34. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **53,** 240–255
35. Pflugrath, J. W. (1999) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55,** 1718–1725
36. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276,** 307–326
37. Collaborative Computing Project 4 (1994) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50,** 760–763
38. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., and Sacchettini, J. C. (2002) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **58,** 1948–1954
39. Desjardins, R. E., Canfield, C. J., Haynes, J. D., and Chulay, J. D. (1979) *Antimicrob. Agents Chemother.* **16,** 710–718
40. Fidock, D. A., Nomura, T., and Wellems, T. E. (1998) *Mol. Pharmacol.* **54,** 1140–1147
41. Nkrumah, L. J., Muhle, R. A., Moura, P. A., Ghosh, P., Hatfull, G. F., Jacobs, W. R., Jr., and Fidock, D. A. (2006) *Nat. Meth.* **3,** 615–621
42. Bernstein, J., Lott, W. A., Steinberg, B. A., and Yale, H. L. (1952) *Am. Rev. Tuber.* **65,** 357–364
43. Fox, H. H. (1952) *Science* **115,** 129–134
44. Banerjee, A., Dubnau, E., Quemard, A., Balasubramanian, V., Um, K. S., Wilson, T., Collins, D., de Lisle, G., and Jacobs, W. R., Jr. (1994) *Science* **263,** 227–230
45. Dessen, A., Quemard, A., Blanchard, J. S., Jacobs, W. R., Jr., and Sacchettini, J. C. (1995) *Science* **267,** 1638–1641
46. Rozwarski, D. A., Grant, G. A., Barton, D. H. R., Jacobs, W. R., Jr., and Sacchettini, J. C. (1998) *Science* **279,** 98–102
47. Vilcheze, C., Wang, F., Arai, M., Hazbon, M. H., Colangeli, R., Kremer, L., Weisbrod, T. R., Alland, D., Sacchettini, J. C., and Jacobs, W. R., Jr. (2006) *Nat. Med.* **12,** 1027–1029
48. Geary, T. G., and Jensen, J. B. (1983) *Am. J. Trop. Med. Hyg.* **32,** 221–225
49. Zhang, Y., Heym, B., Allen, B., Young, D., and Cole, S. (1992) *Nature* **358,** 591–593
50. Nguyen, M., Claparols, C., Bernadou, J., and Meunier, B. (2001) *Chembiochem* **2,** 877–883
51. Rozwarski, D. A., Vilcheze, C., Sugantino, M., Bittman, R., and Sacchettini, J. C. (1999) *J. Biol. Chem.* **274,** 15582–15589
52. Copeland, R. A. (1996) *Enzymes: A Practical Introduction to Structure, Mechanism, and Data Analysis*, pp. 266–349 VCH Publishers, Inc., New York
53. Kapoor, M., Dar, M. J., Surolia, A., and Surolia, N. (2001) *Biochem. Biophys. Res. Commun.* **289,** 832–837
54. Kapoor, M., Mukhi, P. L. S., Surolia, N., Suguna, K., and Surolia, A. (2004) *Biochem. J.* **381,** 725–733
55. Rosenthal, P. J., Wollish, W. S., Palmer, J. T., and Resnick, D. (1991) *J. Clin. Invest.* **88,** 1467–1472
56. Baldwin, J., Michnoff, C. H., Malmquist, N. A., White, J., Roth, M. G., Rathod, P. K., and Phillips, M. A. (2005) *J. Biol. Chem.* **280,** 21847–21853
57. Kumar, S., Kumar, G., Kapoor, M., Surolia, A., and Surolia, N. (2006) *Synth. Commun.* **36,** 215–226
58. Kumar, G., Parasuraman, P., Sharma, S. K., Banerjee, A., Karmodiya, K., Surolia, N., and Surolia, A. (2007) *J. Med. Chem.* **50,** 2665–2675
59. Zhang, Y.-M., and Rock, C. O. (2004) *J. Biol. Chem.* **279,** 30994–31001
60. Kirmizibekmez, H., Calis, I., Perozzo, R., Brun, R., Donmez, A. A., Linden, A., Puedi, P., and Tasdemir, D. (2004) *Planta Med.* **70,** 711–717
61. Tasdemir, D., Lack, G., Brun, R., Ruedi, P., Scapozza, L., and Perozzo, R. (2006) *J. Med. Chem.* **49,** 3345–3353
62. Sharma, S. K., Parasuraman, P., Kumar, G., Surolia, N., and Surolia, A. (2007) *J. Med. Chem.* **50,** 765–775
63. Guex, N., and Peitsch, M. C. (1997) *Electrophoresis* **18,** 2714–2723